Evolution of Transcriptional Regulatory Networks in Pseudomonas aeruginosa During Long Time Growth in Human Hosts

Andresen, Eva Kammer

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Evolution of Transcriptional Regulatory Networks in Pseudomonas aeruginosa During Long Time Growth in Human Hosts

PhD Thesis

by

Eva Kammer Andresen

Department of Systems Biology

Technical University of Denmark

May 2016

DTU Systems Biology
Department of Systems Biology

Preface

Evolution of Transcriptional Regulatory Networks in Pseudomonas aeruginosa During Long Term Growth in Human Hosts

PhD Thesis by
Eva Kammer Andresen

Primary supervisor
Lars Jelsbak, Associate Professor

Co-supervisor
Maher Abou Hachem, Associate Professor

Department of Systems Biology
Technical University of Denmark
Matematik Torvet
Building 301
2800 Kgs. Lyngby
Denmark

Preface

This thesis is written as a partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark (DTU). The work presented in this thesis was carried out from August 2012 to May 2016 at the Infection Microbiology Group and Enzyme and Protein Chemistry Group, Department of Systems Biology at DTU. The work was carried out under the supervision of Associate Professor Lars Jelsbak, Infection Microbiology Group, and Associate Professor Maher Abou Hachem, Enzyme and Protein Chemistry. The work was funded as a PhD stipend from the Technical University of Denmark.

Eva Kammer Andresen

Kongens Lyngby, May 2016

Acknowledgements

I am now at the end of 4 great years. During those years, I have encountered a number of people that each deserves thanks for their help and support in various ways. First of all, I would like to thank my two supervisors, Associate Professor Lars Jelsbak and Associate Professor Maher Abou Hachem. Thanks for time, energy and discussions and for creating a great working environment, whether in the Infection Microbiology Group or in the Enzyme and Protein Chemistry Group.

I would like to thank the entire past and present IMG group, especially Vinoth Wigneswaran, Anne-Mette Christensen, Grith M. M. Hermansen, Cristina Isabel Amador Hierro, Sandra Wingaard Thrane, S. M. Hossein Khademi, Anders Norman, Claus Sternberg, and my officemate Carsten Jers, for a great working environment. Thank to Susanne Kofoed (Søs) for keeping the labs running, what would we do without you? And an extra thank to Grith for jumping in and helping with the last lab work. It was much appreciated! Lea Mette Madsen Sommer, thanks for keeping me company at DTU, keeping me company at various cafes, keeping my pace up on my bike, and for keep writing annoying emails while I was desperately trying to finish this thesis ☺

My work in Enzyme and Protein Chemistry Group has been more intense and troublesome than initially anticipated. A special thanks to Maher for his energy and help, inside as well as outside the lab. I am impressed with the amount of energy and attention you put into each single project. Also thanks to Laura Nekiunaite and Susan Andersen, you definitely made all the long days in 224/375 worthwhile, and I will miss our frustrated discussions on, how to perform and purify from 100L expression systems, as well as why the Avant/Purifier/Explorer is never doing what it should, even though we already pressed all the buttons! ☺

I had the pleasure of visiting the laboratory of Dr. Tino Krell in Granada, Spain for 2 weeks. In spite of the short time, this visit meant a lot to me. It was at a time when I scientifically needed a change of scenery, and I was impressed with the way I was received, and how much help and assistance I received. I want to thank the entire group, and especially Tino for inviting me, for taking out time to teach me ITC, and for truly caring about that I brought home results. Also thanks to Abdelali Daddaoua for teaching me EMSA.

I also had the great pleasure of visiting the laboratory of Prof. Dr. Susanne Haussler at Helmholtz Centre for Infection Research, Braunschweig, Germany. I want to thank Susanne for the chance to

visit her group. Thanks to all the people from the Haussler lab, a few especially; Sebastian Schulz, thank you for suggestions on experimental set-up, and thanks for all the work you did for me in lab. Denitsa Eckweiler, thanks for the hours you have put into data analysis, for your suggestions, and for all the additional, alternative analyses I asked you to do. Also thanks to Tanja Nicolai and Agata Bielecka for great assistance in the lab.

Last, and perhaps most importantly, I need to thank my family. Mads, Malthe, and my mum and dad. My mum and dad are remarkable people who always turn up when we, or anyone, needs help. No matter if it is baby-sitting, vacuum-cleaning, floor-washing, we can count on you. And you even claim to enjoy it! Mads, thanks for being patient and not letting it get to you when I am unreasonable (yes, that does happen!) and thanks for taking care of everything and letting me work long hours when needed. And last but not least, thanks to Malthe, even though he doesn't understand, for keeping my focus on what is really important.

Contents

Contents

Abstract

Bacteria are remarkable organisms with the capacity to adapt to new environments by remodelling their gene expression profiles. The specific genomic material of any bacterium determines its capacity for any gene regulatory repertoire. However, by evolutionary shaping, these regulatory networks are subjected to forces that allow the bacteria to break genomic constraints, remodel existing regulatory networks, and colonise new environments. While experimental evolution studies have documented that global regulators of gene expression are indeed targets for adaptive mutations, it is less clear to which extent these observations relate to natural microbial populations.

The focus of this thesis has been to study how regulatory networks evolve in natural systems. By using a particular infectious disease scenario (human associated persistent airway infections caused by the bacterium *Pseudomonas aeruginosa*) as a natural model system, the work has focused on characterising a number of mutations in global regulators that are known to provide an adaptive advantage in this specific environment. The aim has been to provide a molecular explanation of the effects of the specific mutations in relation to regulatory network remodelling, and to provide insight into the extent of epistasis and evolutionary dynamics of these systems.

The two studies presented in this thesis specifically deal with single amino acid substitutions or deletions in the sigma factors RpoD, AlgT, and RpoN. Through *in vitro* techniques, we characterised the direct molecular effects of the sigma factors' abilities to interact with DNA and the core RNA polymerase (RNAP). By combining this approach with *in vivo* transcription profile data, Chromatin Immunoprecipitation-sequencing (ChIP-seq) data and artificial regulatory network modifications by *in vivo* sigma factor overexpression, we were able to investigate how the altered molecule-to-molecule interactions induce rewiring of transcriptional regulatory networks and create unexpected phenotypes.

The results show that through remodelling of the respective regulatory networks, mutations fixed in global regulator genes facilitate the generation of novel phenotypes which again facilitate the shift in life-style of the bacterium from an environmental opportunistic pathogen to a human airway specific pathogen. These findings are not only applicable to *P. aeruginosa* specific studies, but suggest that, on a general level, evolutionary remodelling of regulatory network structures may be the key to ecological success in the wild.

Dansk resume

Bakterier er bemærkelsesværdige organismer med en ekstraordinær evne til at tilpasse sig fremmede miljøer ved omprogrammering af deres gen ekspression. En bakteries DNA bestemmer kapaciteten for dens gen regulatoriske repertoire. Dog kan evolutionen formgive og ændre gen regulatoriske netværk så bakterier bliver i stand til at bryde arvemasse bestemte restriktioner, omprogrammere de eksisterende regulatoriske netværk og kolonisere nye miljøer. Eksperimentelle evolutions studier har påvist, at globale regulatorer er mål for adaptive mutationer, dog er det debatteret, hvor vidt disse observationer kan relateres til naturlige mikrobielle populationer.

Fokusset i denne afhandling er således at afdække hvordan gen regulatoriske netværk udvikles i naturlige systemer. Ved at anvende kroniske luftvejsinfektioner med den opportunistisk patogene bakterie, *P. aeruginosa* som et naturligt modelsystem, er der fokuseret på at karakterisere specifikke mutationer i globale genetiske regulatorer, som er associeret med adaptation til det specifikke miljø. Formålet har været at afdække de molekylære mekanismer, der resulterer i de effekter som de specifikke mutationer har på omprogrammeringen af regulatoriske netværk, samt at afdække hvorvidt epistasi påvirker dynamikken i disse systemer.

I denne afhandlings to studier er der specifikt blevet arbejdet med aminosyre substitutioner og deletioner i sigma faktorerne RpoD, AlgT og RpoN. Ved hjælp af *in vitro* eksperimentelle teknikker har vi karakteriseret de direkte effekter af mutationerne i forhold til evnen til at interagere med DNA og core RNAP. Ved at kombinere disse teknikker med *in vivo* gen ekspressions studier og Chromatin Immunoprecipitation-sequencing (ChIP-seq), samt artificielle modifikationer af regulatoriske netværk ved hjælp af *in vivo* sigma factor overekspression, har vi været i stand til at undersøge hvordan de ændrede molekyle-molekyle interaktioner inducerer en omprogrammering af gen regulatoriske netværk og resulterer i uforudsete fænotyper.

Resultaterne viser, at gennem omprogrammering af de respektive gen regulatoriske netværk kan mutationer der er fixeret i globale genetiske regulatorer afstedkomme nye fænotyper. De nye fænotyper muliggør et skift i bakteriens livsstil fra værende en miljø-associeret opportunistisk patogen, til en luftvejsspecifik patogen der er optimeret til det nye miljø. Disse resultater er ikke kun interessante i forhold til andre specifikke studier af *P. aeruginosa* men kan bruges til at forstå hvordan en evolutionær drevet omprogrammering af regulatoriske netværk baner vejen for en organismes succes i et specifikt miljø.

Abbreviations

bEBP	Bacterial Enhancer Binding Protein
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Regulator
ChIP-seq	Chromatin Immunoprecipitation-sequencing
DIP-seq	DNA Immunoprecipitation-sequencing
ECF	Extracytoplasmic Functioning
EMSA	Electrophoretic Mobility Shift Assay
HTH	Helix-turn-helix
IHF	Integration Host Factor
NGS	Next generation sequencing
PMNs	Polymorphnuclear leukocytes
ppGpp	Guanosine tetra- and pentaphosphate
QS	Qurom Sensing
RNAP	RNA polymerase
ROS	Reactive Oxygen Species
SNP	Single Nucleotide Polymorphism
SPR	Surface Plasmon Resonance
tad-locus	Tight adherence locus
TF	Transcription Factor
TRN	Transcriptional Regulatory Network
TSS	Transcription Start Site
UAS	Upstream Activating Sequence

List of figures

List of publications

Paper 1

Eva Kammer Andresen, Denitsa Eckweiler, Sebastian Schulz, Susanne Haussler, Tino Krell Maher Abou Hachem, Lars Jelsbak (2016). **Rewiring of a sigma factor regulatory network in *P. aeruginosa* by a naturally occurring single nucleotide polymorphism**. Manuscript submitted for publication.

Paper 2

Eva Kammer Andresen, Denitsa Eckweiler, Sebastian Schulz, Grith Mirriam Maigaard Hermansen, Susanne Haussler, Maher Abou Hachem, Lars Jelsbak (2016). **Epistasis and sigma factor competition in *Pseudomonas aeruginosa.*** Manuscript in preperation.

“Nature does not hurry, yet everything is accomplished”

Lao-Tzu, Chinese Philosopher, 6th–5th century BC

Chapter 1

Introduction and thesis overview

Ever since the beginning of life, organisms have evolved and diversified. A constant need to adapt to new and changing environments, as well as a need for responding to interacting species has resulted in the ongoing process of evolution that we with modern-day techniques and computing power can visualise and begin to unravel at much greater depth than just 15 years ago. One of the major goals in the field of evolutionary biology is to understand the evolutionary dynamics and genetic basis of adaptation. However, this field is not of an isolated interest for evolutionary biologists only. The greater technological advances we as humans produce, the more clear it becomes not only that a molecular and mechanistic understanding of evolution is central for our ability to comprehend the development of life on earth, but also that a molecular and mechanistic understanding of evolution will be the prerequisite of many new technological advances, such as understanding disease and infection scenarios, the possible and intimidating future use of human genetic manipulations, or industrial use of cell factories. This thesis specifically deals with the evolution of bacterial transcriptional regulatory networks, though most of the problems that will be touched upon relate to all species. The diversity and adaptability of bacteria is of great value in industrial applications. The same features, however, make them formidable disease causing agents that possess great potential of evading medical treatment. An in-depth understanding of the evolution and molecular interplay is therefore important for our ability to engineer and control these organisms.

The focus of this thesis has been to increase our understanding of how bacterial transcriptional regulatory networks (TRNs) evolve, and how different molecular mechanisms co-operate to produce the most adapted phenotype for a given environment. The model system used for this investigation has been long-term chronic airway infections in Cystic Fibrosis (CF) patients caused by the opportunistic pathogen *P. aeruginosa,* and while the findings in this thesis may be specific

for *P. aeruginosa* in the CF lung environment, the problems presented relate to all fields of microbial evolution and genetic adaptation, and are therefore of general interest to the microbiological society.

1.1 Thesis outline

This thesis is organised into seven chapters. While the current chapter (**Chapter 1**) introduces the thesis, **Chapter 2** presents a general introduction to evolution and adaptation. This includes an introduction to varies forms of mutations that drive evolution, as well as evolution studies of both experimental and natural systems. **Chapter 3** gives a more in-depth introduction to bacterial gene regulation, with special focus on the process of transcription, involving sigma factors, sigma factor competition, and factors that influence sigma factor competition. **Chapter 4** introduces the concept of TRNs and their evolution, in particular related to the evolution and adaptation of *P. aeruginosa* to the CF lung environment. **Chapter 5** provides an introduction to the investigations presented in this thesis, as well as the background and the thesis aim. **Chapter 6** provides an overall conclusion and perspective of the thesis, and finally, **Chapter 7** contains the research articles presented in this thesis in full length.

Chapter 2

Microbial evolution

Bacteria are incredible diverse organisms, with a fascinating capability to survive in even the most harsh and extreme environments. From hyperthermophiles thriving in high temperatures >80°C and halophiles thriving in high salt concentrations (2-5 M NaCl), to alkaliphiles and acidophiles thriving in high and low pH and even the bacterium *Deinococcus radiodurans* which tolerates radiation (Rothschild & Mancinelli 2001), it is incredible that these organisms, all build from the same building blocks as any other organism, are capable of tolerating such extreme environments that would immediately put an end to most other lifeforms. Bacteria are small organisms, not able to physically move great distances or at high speed. So how then, do they adapt to and colonise these extreme environments that no other organism is capable of enduring?

Bacteria have an extraordinary ability to evolve and adapt to novel environments. Their small genome sizes and short life cycles enable them to reproduce fast, thus allowing evolutionary modification to settle at a pace that far exceeds e.g. the human genetic evolution. Two different mechanisms account for the ability to adapt to novel environments; phenotypic adaptation and genotypic evolution. Phenotypic adaptation involves alterations of gene regulation, and a resulting altered phenotype, but with no inheritable genetic changes. Phenotypic adaptation is thus a response that can be turned on, if needed, and is not inherited to later descendants of the population (Brooks et al. 2011). Phenotypic adaptation in the form of gene regulation will be discussed in depth in Chapter 3. Phenotypic adaptation is a quick response to short-term fluctuations in the environments, however, has limited capacity and may be inadequate during long-term adaptation to novel environments. In such cases, the organism is dependent on genotypic evolution, which involves genetic changes, that are inherited to descendants, thus resulting in an alteration of the population structure and a selection in favour of the individuals that are capable of producing the most beneficial phenotype for the given situation (Brooks et al. 2011).

The following sections will introduce the reader to distinct types of genetic adaptation, as well as environmental forces driving genotypic adaptation, and how evolutionary biologists can study bacterial genetic evolution in natural and artificial systems.

2.1 Microbial genetic evolution

With the advent of whole-genome sequencing researchers now have the tools to identify the underlying genetic steps of evolution, rather than being confined to the limitation of observing the evolved phenotypes. Genotypic evolution is any kind of genetic alteration or rearrangement that is inherited in the population. Point mutations in the existing genetic material, rearrangement of the existing genetic material, gene duplication events, acquisition of new genetic material, or deletion of genetic material all account for genetic evolution, though each kind facilitates evolution through a different mechanism and at different pace (Koskella & Vos 2015). Acquisition of new genetic material provides the bacteria with a fast mechanism of adapting to new environments through one single step by the uptake of genetic material that encodes for cellular functions encoded by several genes (Wiedenbeck & Cohan 2011). The large-scale insertions of genetic material often includes addition of genes coding for antibiotic resistance or virulence factors and is thought to play a major role in the emergence of epidemic strains and new species (Bryant et al. 2012).

Acquisition of genetic material in the form of pathogenicity islands was for a long time thought to be the major mechanism of pathogen evolution, and while transfer of large sections of DNA indeed may result in evolution of pathogenicity, and emergence of pandemic outbreaks (Bryant et al. 2012), it has become clear that evolution of especially bacterial pathogens are also largely driven by genomic deletion events (Merhej et al. 2013). Comparative genomics has shown, that during the transition from a free-living lifestyle to a host-associated pathogen, gene loss confer a fitness advantage, as well as loss of transcriptional regulators and disruption of ribosomal RNA operons seem to trigger the development of pathogenicity (Merhej et al. 2013)

Deletion of genomic content has indeed been shown to be fitness-increasing in a population of *Salmonella Enterica* grown in rich media in a laboratory experiment (Koskiniemi et al. 2012), and also in natural systems, exemplified by the *P. aeruginosa* DK2 clone adapting to the Cystic Fibrosis lung environment, deletion of genetic material has shown to be driving factor for host adaptation, rather than acquisition of new DNA (Rau et al. 2012). Especially the early stages of *P. aeruginosa* adaptation to the CF airway were characterised by loss of genomic material, with an overrepresentation of reduction of the accessory genome.

2.1.1 Point mutations

While genomic deletions and acquisitions may produce instant additions or deletions of entire cellular functions during evolution, evolution in the form of point mutations is equally important, despite their relative small changes compared to the total genetic material.

Point mutations in the existing genetic material is an evolutionary slow process and generally occur at low frequencies in the bacterial genome with an estimated mutation rate as low as 10^{-10}-10^{-9} per base pair per replication (Barrick & Lenski 2013). Point mutations occur either as a substitution of one nucleotide with another (referred to as a single nucleotide polymorphism (SNP)), or as an insertion or deletion of a nucleotide (Bryant et al. 2012). Traditionally, point mutations have been difficult to study, as they require advanced sequencing technology to detect, and analyses that compare and detect point mutations among bacterial species and clones require both skill and computing power. However, the constant advances of Next Generation Sequencing (NGS) techniques have enabled detailed studies of all aspects of microbial evolution.

Though point mutations may seem as small changes providing incremental evolutionary steps, evidence from recent studies point to, that evolution of pathogenic potential and host adaptation often involves rewiring of pre-existing regulatory networks in the pathogen genome, and that this rewiring is largely driven by non-synonymous mutations in global regulator genes (Renzoni et al. 2011; Damkiaer et al. 2013; Flores et al. 2015). For example, it has been shown that in the group A streptococci, sequence variation in the control of virulence regulator (CovR) protein directly affects the transcriptome and virulence profile (Horstmann et al. 2011), and that a mutation in the sensor kinase LiaS results in an alteration, but not elimination of the LiaS protein function (Flores et al. 2015).

Often, the phenotypic consequences of emerging SNPs are enforced by epistasis, which is defined as interactions between two mutations at different loci that produce an effect on a phenotype that deviates from the sum of their individual effects (Elena & Lenski 2003). This was exemplified by Damkiaer et al. 2013, who investigated the phenotypic consequences of specific global regulator mutations, fixed in the *P. aeruginosa* DK2 lineage during evolution and adaptation to the CF lung environment. The study showed that only few mutations in global regulator genes are necessary to reflect the phenotypic transition from an opportunistic pathogen to a primary host-specific pathogen.

2.2 Studying microbial evolution – evolution experiments and natural model systems

Historically, studying bacterial evolution has been difficult, partly due to technical difficulties that arise from the microscopic sizes of these organisms. Recent technological advances have not only enabled the study of bacterial evolution, it has also highlighted the importance of understanding how bacteria evolve and adapt to new environments, such as in disease scenarios (Koskella & Vos 2015). With recent years´ technological advances, microorganisms have increasingly been used for evolution studies. Not only do microorganisms represent a field of special interest – microorganisms are important players in everything from industry to health and disease. They also represent the perfect organism for evolution studies. They reproduce quickly and asexually, they are easily grown and stored, and can be kept as frozen fossil records ready to be resurrected for further testing. Their genetic material is easily manipulated and with the technological advances during recent years, genomic, proteomic, transcriptomic, and a number of other high throughput techniques allow for a detailed mapping of not only genetic changes, but also downstream responses to these changes (Barrick & Lenski 2013).

2.2.1 Evolution experiments

In an evolution experiment, a population of microorganisms is established and propagated in a controlled environment, and ancestral samples, as well as samples from different timepoints of the experiments are taken and stored for subsequent analysis (Koskella & Vos 2015). By storing samples, the evolution rate and phenotypic consequences can be measured directly, e.g. by comparing fitness/growth rates of the evolved strain to the ancestor, and this can be directly linked to the genotypic evolution from sequencing of the ancestral and evolved genomes. While evolution experiments may present an artificial system that lack the complexity of most natural systems, they are suitable for answering fundamental evolutionary questions such as whether genetic adaptation continues indefinitely, even in a constant environment, how individual mutations contributes to fitness improvement, as well as how reproducible evolutionary changes are (Elena & Lenski 2003).

The longest running experimental evolution study, initiated in 1988, studied the evolution of *Escherichia coli* during continuous serial propagations (Philippe et al. 2007). The evolved populations from this experimental evolution study have shown that parallel phenotypic changes, genetic parallelism and rewiring of biological networks are accessible on the level of structural genes, as well as global regulator genes (Hindré et al. 2012).

While evolution experiments have provided numerous important findings about genomic evolution and adaptation rates, how beneficial and neutral mutations accumulate, and also how small-scale specific genomic rearrangement can create extensive regulatory and metabolic remodelling that facilitates entirely new cellular capacities (Barrick et al. 2009; Blount et al. 2012) there are certain drawbacks from studying evolution in controlled experiments. As implied by the name, parameters such as population size, interacting organisms, nutrients, etc. are controlled (Elena & Lenski 2003). While this may be advantageous in certain ways, is also raises the question about, to which extent findings generated from evolution experiments are comparable to natural systems that exhibit a much greater variation in nutrients, interacting organisms, as well as includes sub-niches which allow for niche differentiation

2.2.2 Cystic Fibrosis as a natural model system

The CF lung environment represents a biologically relevant model system for which studying microbial evolution produces both relevant basic knowledge about microbial evolution, but also contributes important knowledge on infectious disease scenarios that may represent a significant improvement to the patients' health; an improvement that may be expanded to other infectious diseases. CF is a genetic autosomal recessive disorder, affecting about 1 in 2500 individuals (Bye et al. 1994). In individuals suffering from CF, mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene causes a defective chloride ion transport across epithelial cell surfaces, which leads to extremely viscous lung mucus (Welsh & Smith 1993). Whereas healthy individuals remove inhaled microorganisms from their airways by coughing, the viscous lung mucus of the CF lung impairs airway clearance, thus allowing microorganisms to colonise and adapt to the lung environment (Cutting 2014).

The CF lung environment

The CF lung is a highly heterogeneous environment and presents a number of stressful conditions for colonising bacteria. The dynamics of this system changes in both time and space, and forces incoming microorganisms to constantly adapt to these varying factors. Not only does the natural physiological environment present an environmental shift for the bacteria, such as a heterologous distribution of oxygen, salts, and nutrients, in addition they are constantly combatted by polymorph nuclear leukocytes (PMNs) and aggressive antibiotic treatment as well as stressed by reactive oxygen species and reactive nitrogen intermediates produced by the PMNs (Yang, Jelsbak & Molin 2011; Folkesson et al. 2012). Compared to controlled evolution experiments, the CF lung environment is not kept constant, and nutrients as well as stress factors vary over time, creating a

constant need for a flexible adaptation that must accommodate both metabolic adaptation to the shift in nutrient distribution, adaptation to survive the aggressive antibiotic treatment as well as adaptation to survive, conquer or benefit from the interaction with other microbial species (Sousa & Pereira 2014).

Microbiology of the CF lung

The most frequent colonisers of the cystic fibrosis lung are *P. aeruginosa*, *Haemophilus influenza*, and *Staphylococcus aureus* with *P. aeruginosa* as the dominant pathogen present in about 80% of adult CF patients (Harrison 2007). Microorganisms entering the lungs initially establish an intermittent colonisation of the airways. This colonisation can be combated with aggressive antibiotic treatment, as well as reoccurring eradication and emergence of bacterial strains. However, intermittent colonisation always transitions into a chronic infection, originating from bacterial seeding from the sinuses where the bacteria have survived and adapted to their new environment (Figure 1) (Folkesson et al. 2012). This chronic infection state is characterised by continuous presence of *P. aeruginosa*, chronic inflammation and PMNs that lead to respiratory failure with extensive airway destruction (Folkesson et al. 2012).

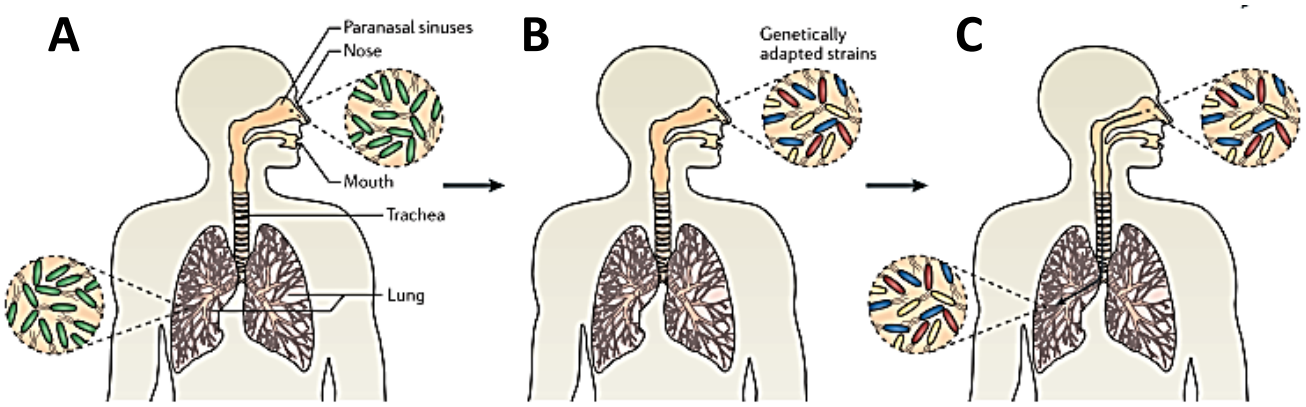


Figure 1. Generalised course of the CF airway colonisation and infection with *P. aeruginosa*. (A) An environmental strain of *P. aeruginosa* invades the CF airways and initiates colonisation of both lungs and sinuses. (B) During aggressive antibiotic treatment as well as attack from the host immune responses, the lung colonisation is eradicated. However, the sinuses provide a more protected environment for the bacteria, where they are thought to be able to colonise and adapt to the new environment. (C) At a later stage, the bacteria from the sinuses, now adapted to the new environment, are able to seed the lungs and establish a chronic infection. Adapted from (Folkesson et al. 2012).

2.2.3 Evolution of *P.* aeruginosa in the CF airway

As mentioned, adaptation of *P. aeruginosa* to the CF lung involves interaction with a more varied and interchangeable environment than those present in controlled laboratory experiments, and clinical isolates sampled from the same patient at the same time exhibit phenotypic variety (Mowat et al. 2011). Even though the extent of phenotype diversity varies between patients, a number of specific phenotypes are repeatedly observed in clinical sample from the CF lung. These CF evolved phenotypes are as mentioned a product of the specific environments that exert an evolutionary force on the bacteria.

A common *P. aeruginosa* CF evolved phenotype is the mucoid phenotype. The mucoid phenotype is a result of an overproduction of the exopolysaccharide, alginate, which results in a biofilm-like colony morphology of *P. aeruginosa*. The overproduction of alginate protects *P. aeruginosa* from a number of factors such as inflammatory effects, decreased phagocytosis by PMNs and macrophages, and reactive oxygen species (ROS) (Lyczak et al. 2000). The mucoid phenotype is unstable in vitro, and is often observed to revert back to a nonmucoid phenotype, suggesting that this is a specific CF evolved trait (Ciofu et al. 2001). The genetic machinery required for alginate production is an inherent feature of the *P. aeruginosa* genomic potential, and switches in this phenotype often arises from point mutations that remodel existing regulatory networks. Regulation of alginate production and the emergence of the mucoid phenotype is a recurring subject in this thesis, and a more in-depth discussion of the genetic basis for alginate production is provided in Chapters 3 and 4.

Other common CF evolved traits are the development of antibiotic resistance and loss of virulence factors. Antibiotic treatment is used extensively during infections of CF patients, and *P. aeruginosa* is intrinsically resistant towards many antibiotics due to low outer membrane permeability, and the presence of efflux pumps, and mutations in regulators of efflux pumps typically lead to their overexpression, causing high levels of antibiotic resistance (Poole 2001). Loss of virulence factors includes flagella, type IV pili, as well as a number of secreted factors such as proteases, siderophores and factors of the type III secretion system. Loss of virulence factors arise through mutations affecting global regulators such as a number of sigma factors and the quorum-sensing regulators, Vfr and LasR (Smith et al. 2006; Yang, Jelsbak, Marvig, et al. 2011). With the large genetic repertoire of *P. aeruginosa*, it is clear that much of the evolution of this bacterium in the CF environment is dependent on an extensive remodelling of regulatory networks and gene expression

levels, and that only few mutations in global gene regulators cause large disruptions of the transcriptional regulatory networks with resulting epistatic effects (Damkiaer et al. 2013).

Chapter 3

Bacterial gene regulation

Bacterial gene regulation is the process ensuring a bacterium to express the appropriate genes, at the right time, at the exact needed amount. At any time, an immense number of regulatory factors secure that each cell produce only the absolutely needed number and kind of mRNA and proteins. Being the omnipresent machinery of the cell, it is obvious that the nature and ways of genetic regulation must be as complex and extraordinary as bacteria themselves.

The types of regulatory factors controlling gene regulation and expression in bacteria seem never-ending. Proteins such as sigma factors, transcription factors (TFs), and two-component systems (TCS), DNA, non-coding RNA, small inhibitory RNA, micro RNA, as well as a number of signal molecules such as guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp), collectively referred to as ppGpp herein, all assist in controlling cellular processes. While it is far beyond the aim of this thesis to introduce a detailed description of the entire genetic regulatory machinery, the following sections will provide an overview of general bacterial gene regulation and focus in detail on a smaller subset of genetic regulators, those´ whose functions have been studied in the appended research papers, and those whose´ function is directly related to the process of gene transcription.

3.1 Transcription

Transcription is the process of transcribing DNA to RNA. The process is initiated by the DNA directed RNA polymerase, RNAP. The RNAP is a large enzyme consisting of 6 subunits ($\alpha_2\beta\beta'\omega\sigma$) and capable of initiating transcription from promoter sequences. The catalytic machinery capable of polymerising long RNA chains resides within the $\beta\beta'\alpha_2\omega$ complex, whereas the σ subunit provides translational direction to the RNAP enzyme, directing it to specific promoters of genes that need expression (Burgess & Travers 1969). A RNA polymerase complex devoid of the σ factor subunit is denoted as the core RNAP, whereas the core RNAP in complex with the σ factor is denoted holo RNAP.

Transcription from a gene is initiated when the core RNAP associates with the σ factor to form the holo RNAP (Figure 2.a) and thereafter recognises and binds to the promoter sequence (Figure 2.b).

After binding to the promoter in this stable closed complex, the holo RNAP forms an open complex by separating the double stranded DNA and initiates elongation (Figure 2.c-e). The growing RNA strand produces unfavourable kinetics constrains on the holo RNAP, and the sigma factor is released (Figure 2.f) (Gill et al. 1991). The core RNAP is thus able to slide along the DNA while producing a growing RNA chain. After release, the sigma factor is again able to associate with a new core RNAP and directing transcription from a new promoter, a process known as the sigma cycle (Mooney et al. 2005; Raffaelle et al. 2005).

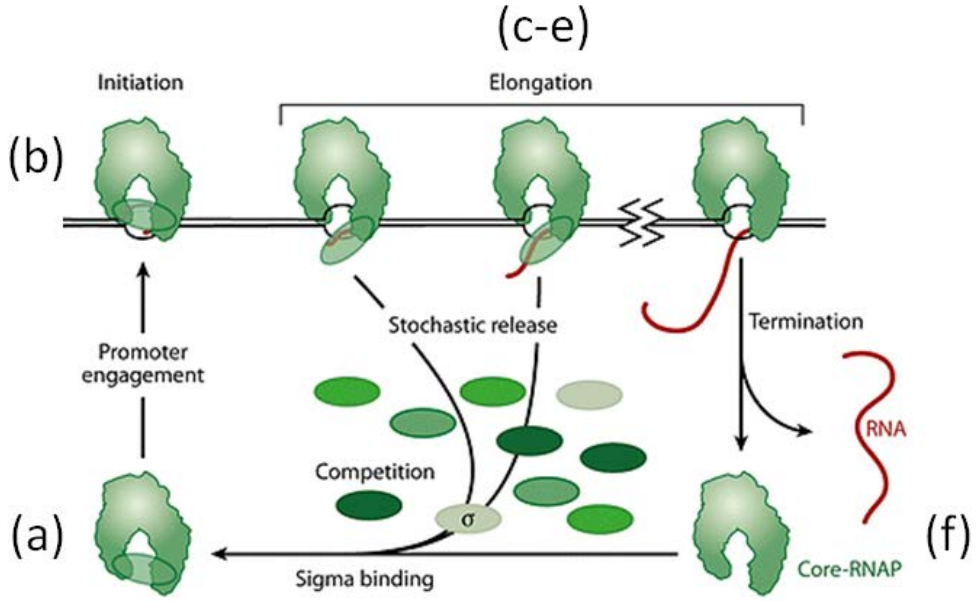


Figure 2. **Representation of bacterial transcription initiation by the holo RNAP complex and the sigma cycle.** Transcription is initiated when a sigma factor from the cellular pool gains access to, and binds the core RNAP (a). The resulting holo RNAP then recognises and binds a sigma-specific promoter sequence (b). The holo RNAP then forms an open complex by separating the double stranded DNA and initiates elongation (c-e). The growing RNA strand produces unfavourable kinetics constrains on the holo RNAP, and the sigma factor is released (f), after which the core RNAP is again available for complex formation to a sigma factor. Modified from (Österberg et al. 2011).

3.1.1 Sigma factors

While the 5 subunits forming the core RNAP are highly conserved among bacteria and exist as a single type, the number of different types of σ subunits varies greatly between bacterial species,

from only 1 in *Mycoplasma* sp. to an astonishing 109 sigma factors in the Gram-negative myxobacterium *Sorangium cellulosum* (Gruber & Gross 2003; Han et al. 2013). Sigma factors are multi-domain proteins that, based on sequence alignment, can be divided into 2 major families, σ^{70} and σ^{54}, which display little sequence conservation (Merrick et al. 1987).

σ^{70} family structure and mechanism

The σ^{70} family is a broad family that, based on sequence homologies, can be further subdivided into 4 groups (Lonetto et al. 1992). While all members of this family recognise the classic core recognition -35/-10 promoter element, the different specific consensus sequences are the core determinant of each sigma factor´s promoter specificity and ability to direct transcription from specific genes (Helmann & Chamberlin 1988).

Group I represents the primary sigma factors involved in housekeeping functions and indispensable for growth. The housekeeping sigma factor, σ^{70}, (in *E. coli* and *P. aeruginosa*, this enzyme is denoted RpoD), belongs to group I. Group II comprises sigma factors that are dispensable for growth, but still closely related to group I sigma factors and is represented by the sigma factor regulating cellular responses to stress, RpoS. Group III sigma factors control cellular processes such as responses to heat shock, sporulation and flagellar biosynthesis. The last group, group IV sigma factors, are sigma factors that are more distantly related in terms of sequence analysis as they lack some regions present in group I-III sigma factors. Sigma factors from group IV, also known as extracytoplasmic functioning sigma factors (ECF), are involved in controlling responses to cellular stimuli and membrane functions. Group IV sigma factors represent a stripped down version of group I-III sigma factors and contain only regions 2 and 4, whereas group I-III contain 4 structural domains (Figure 3) (Lonetto et al. 1992; Gruber & Gross 2003; Österberg et al. 2011).

Figure 3 shows an illustration of the domain organisation of sigma factors belonging to the σ^{70} family. Region 1 is a less conserved region, with an auto inhibitory domain that inhibits DNA binding in free σ^{70} (Gruber & Gross 2003). Region 2 is subdivided into 4 regions each involved in functions including binding to the core RNAP (region 2.1), melting of DNA (Region 2.3), and recognition of the -10 and -35 promoter element (region 2.4 and region 4.2) (Lesley & Burgess 1989; Gruber et al. 2001; Burgess & Anthony 2001; Gruber & Gross 2003).

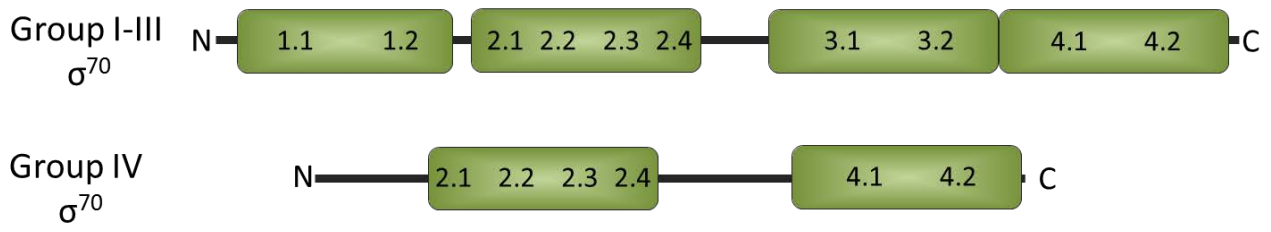


Figure 3. Schematic representation of the regions of sigma factors belonging to the σ^{70} family. (A) Representation of the structural domains composing group I-III sigma factors with regions 1-4 as well as linker regions. (B) Representation of the structural domains composing group IV sigma factors with only regions 2 and 4. Inspiration from (Gruber & Gross 2003; Österberg et al. 2011).

The role of region 3 is less well defined. The region is divided into two subregions, 3.1 and 3.2, which have been appointed several functions such as the binding of initiating NTPs, RNA priming, and promoter recognition, opening, and escape (Severinov et al. 1994; Kulbachinskiy & Mustaev 2006; Pupov et al. 2014). Specific amino acid substitutions of region 3.2 have also been linked to the suppression of growth defects of an *E. coli* ppGpp° strain, as well as being involved in sigma affinity for the core RNAP (Zhou et al. 1992; Hernandez & Cashel 1995; Cashel et al. 2003).

Besides being a major determinant for core binding (region 4.2), region 4 is involved in binding to the -35 promoter region (region 4.2) (Sharp et al. 1999; Burgess & Anthony 2001).

σ^{54} family structure and mechanism

The σ^{54} family consists only of the σ^{54} protein denoted RpoN in *E. coli* and *P. aeruginosa*. Little sequence homology exists between the σ^{54} family and the σ^{70} family (Merrick et al. 1987), and their regulatory mechanisms differ from those of the σ^{70} family members in regard to both promoter recognition and the mechanism of transcription initiation.

The σ^{54} sigma factors recognise consensus promoter signatures consisting of conserved GG and GC residues at the -24/-12 position (Taylor et al. 1996; Buck et al. 2000), and while the σ^{70} family sigma factors are not able to bind DNA without initial holo RNAP complex formation, the σ^{54} sigma factors are able to directly bind to promotor DNA (Buck & Cannon 1992). After DNA binding, the σ^{54} –DNA complex associates with the core to form the holo RNAP in an inactive, closed complex. Transition to an open, active complex is facilitated by binding of a bacterial Enhancer Binding Protein (bEBP) that via ATP hydrolysis creates an open complex is able to

initiate transcription (Studholme & Dixon 2003). bEBPs bind 80-150 nucleotides upstream the -12/-24 promoter sequence, and require an Integration Host Factor (IHF) to bind DNA, which facilitates a loop that brings the bEBP in physical contact with the inactive σ^{54} holo–DNA complex, illustrated in Figure 4 (Shingler 2011).

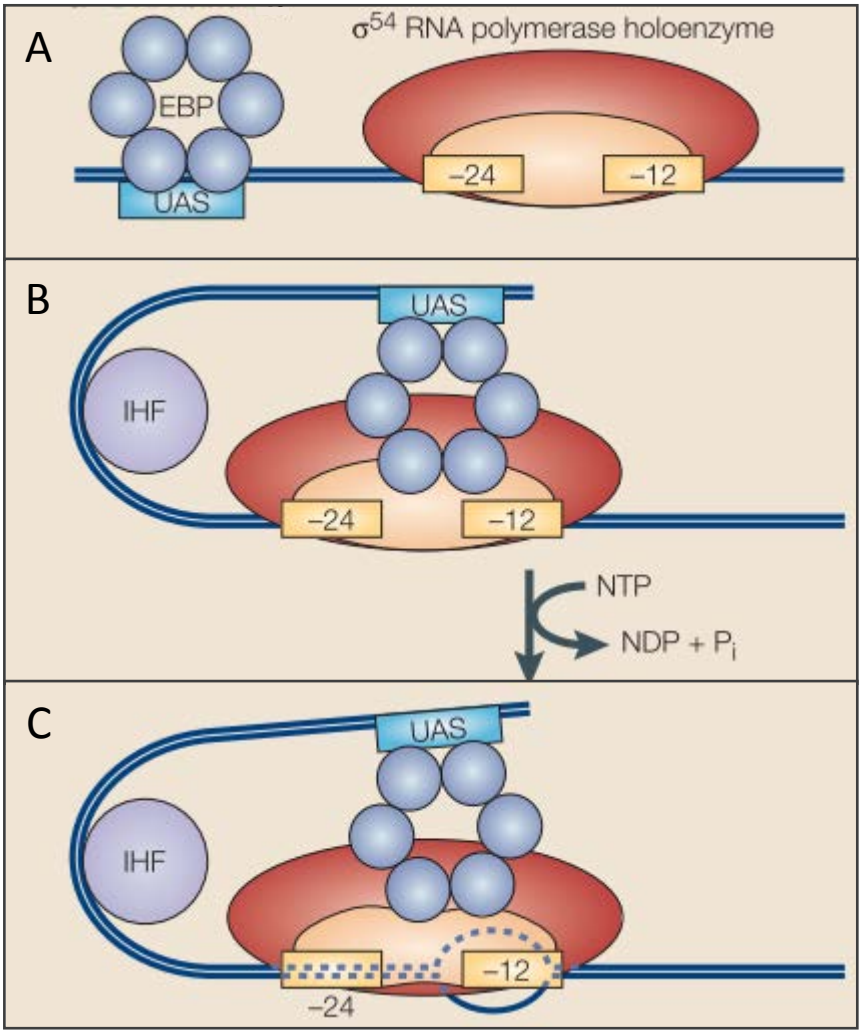


Figure 4. Regulatory mechanisms of σ^{54} dependent transcription at -12/-24 promoter sequences. A: The σ^{54} binds the DNA the promoter sequence and forms the σ^{54} holo-DNA complex. 80-150 nucleotides downstream this site, the specific bEBP binds to an upstream activating sequence (UAS). B: An IHF binds DNA and facilitates DNA looping, which brings the bEBP in physical contact with the σ^{54} holo-DNA complex and facilitates ATP hydrolysis. C: Finally, the ATP hydrolysis favours open complex formation and transcription is initiated. Modified from (Dixon & Kahn 2004)

The structure of the σ^{54} family is composed of three structural domains (Figure 5). Region 1 is involved in interaction with bEBPs, as well as inhibition of DNA binding. Region 2 is a less conserved region and variable in size in different organisms. This linker region is flexible and relocates during RNA synthesis. Region 3 consists of a domain involved in core RNAP binding, a helix-turn-helix (HTH) motif involved in interaction with the -12 promoter region, as well as the RpoN box, involved in interactions with the -24 promoter sequence (Yang et al. 2015).

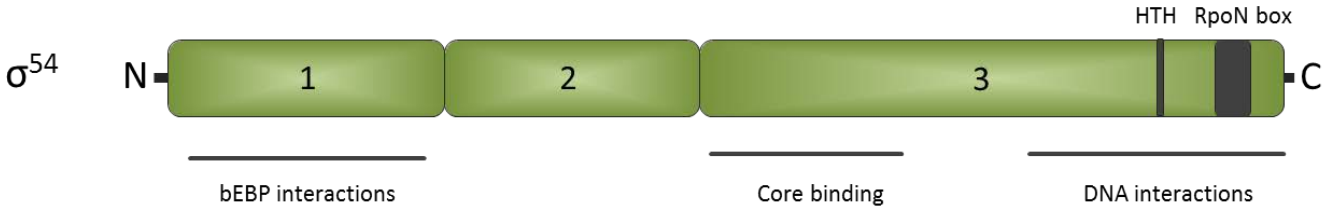


Figure 5. **Schematic representation of the three regions composing σ^{54}.** Regions involved in bEBP binding, core interaction, and DNA binding are marked above and below the figure. (Österberg et al. 2011; Yang et al. 2015).

3.2 Regulation of sigma factor activity

The regulatory potential stemming from the vast numbers and classes of sigma factors alone constitutes a great potential for balancing regulatory responses to environmental fluctuations. But how, then, does the cell ensure proper control between the numbers and kinds of sigma factors that form holo RNAP and thus activate transcription? The following sections dedicated to descriptions of systems and mechanisms that ensure just this: To make sure that only the exact number and kind of sigma factors needed to perform a distinct transcriptional activity is allowed access to the core RNAP.

3.2.1 Sigma factor competition

The reversible binding of sigma factors to the core RNA polymerase allows for competition between the different sigma factors for the core RNA polymerase. While association of the housekeeping sigma factor for the core RNAP is favoured by the numbers of housekeeping sigma compared to alternative sigma factors, bacteria modulate the activity of the alternative sigma factors through several strategies such as anti-sigma factors, 6S RNA, and a range of transcriptional regulators (e.g. activators or inhibitors). By shifting the outcome of sigma factor competition for the

core, these modulators help to redirect the transcriptional profile of the cell to accommodate a given situation (Mauri & Klumpp 2014).

The model of sigma factor competition has been supported by both *in vitro* and *in vivo* experiments which find that both altered concentrations of different sigma factors or reduced binding affinities to the core RNAP modulate the transcription profile. For example, overexpression of one sigma factor results in the downregulation of genes controlled by another sigma factor (Hicks & Grossman 1996; Farewell et al. 1998; Yin et al. 2013), and several cases of mutations in sigma factor encoding genes have reported that mutations decrease the affinity to the core RNAP, thereby lowering transcription activity from genes controlled by the specific sigma factor (Zhou et al. 1992; Zhou & Gross 1992). Furthermore, the relative binding affinities between 6 *E. coli* sigma factor subunits to the *E. coli* core RNAP were determined, and it was shown that incubating a sigma factor subunit with a relatively higher affinity for the core RNAP with a sigma factor with lower affinity for the core RNAP did not result in displacement of the high-affinity sigma factor (Maeda et al. 2000).

Sigma factor competition, and therefore, the transcriptional profile of the cell is as mentioned dependent on both the affinities of a sigma factor to the core RNAP, as well as the actual numbers of molecules present in a cell at a given time. Measuring binding affinities between two proteins is at least conceptually simple, and the absolute affinities between sigma factor subunits and the core RNAP have been determined in a number of studies (Ferguson et al. 2000; Maeda et al. 2000; Rollenhagen et al. 2003; Ganguly & Chatterji 2012). It is clear from those studies that while each sigma factor subunit varies in the affinity to the core RNAP, the experimental conditions such as temperature and salt concentrations are the most important parameters in these studies, and estimating the exact numbers of any given sigma factor and core RNAP that is biologically available for competition is an even more complicated task.

For example, the numbers of core RNAP available to interact with a certain sigma factor are realistically a fraction of the total number of RNAP, because a minor number of core RNAP is already bound to a sigma factor, and a large number of core RNAP is unavailable for sigma competition as they are already engaged in transcription elongation (Mauri & Klumpp 2014). Despite disagreement on the total numbers, most results seem to be in agreement that the total number of sigma factor molecules exceed that of RNAP, thus creating competition among sigma factor subunits for binding to the core RNAP (Laurie et al. 2003; Nyström 2004; Grigorova et al. 2006; Mauri & Klumpp 2014).

3.2.2 Factors influencing sigma factor competition

Anti-sigma factors

While sigma factors control the activation and activity of certain promoters, their own activity is controlled by a group of regulatory proteins referred to as anti-sigma factors, by a mechanism that ensures a fast and robust transcriptional control to sudden environmental changes. Anti-sigma factors exist either as free, cytoplasmic proteins, or as membrane bound, extra-cytoplasmic functioning anti-sigma factors (Hughes & Mathee 1998). Free, cytoplasmic anti-sigma factors function by directly binding the sigma factor at the core-interacting regions, thereby preventing formation of the holo RNAP. An example of this is the sequestering of the *E. coli* RpoD by its anti-sigma factor Rsd through interaction with both RpoD region 2 and region 4 (Dove & Hochschild 2001). An Rsd orthologue, AlgQ, has been identified in *P. aeruginosa*. Activity of AlgQ is required for activation of the genes involved in alginate biosynthesis (Deretic & Konyecsni 1989), and the knowledge that AlgQ interacts directly with *P. aeruginosa* RpoD provides a direct link for transcriptional control by anti-sigma factors (Dove & Hochschild 2001).

Membrane bound, extra-cytoplasmic functioning anti-sigma factors exercise their inhibitory function differently than those of the free cytoplasmic anti-sigma factors. The membrane bound anti-sigma factors control the activity of ECF sigma factors and function via a mechanism known as regulated intermembrane proteolysis (RIP), and contain a periplasmic sensor domain and a cytoplasmic sigma factor binding domain (Hughes & Mathee 1998; Paget 2015). In *P. aeruginosa* this is exemplified by inhibition of the ECF sigma factor AlgT by the anti-sigma factor MucA and its complex forming partner, MucB. MucA spans the inner membrane of *P. aeruginosa*. Its C-terminal domain in the periplasmic space binds another protein, MucB, which protects MucA from proteolytic degradation and sensor protein. The N-terminal of MucA binds AlgT, thus inhibiting it from interacting with the core RNAP (Schurr et al. 1996; Mathee et al. 1997; Cezairliyan & Sauer 2009). Stress signals, such as incorrectly folded proteins in the membrane leads to proteolytic cleavage of MucA and MucB, thereby release of AlgT and activation of the AlgT regulon (Damron & Goldberg 2012). The activity of the Anti-sigma factors themselves is regulated by a variety of mechanisms, spanning from proteolytic degradation, as in the case of MucA, to export out of the cell. A complete description of the regulation of the anti-sigma factors themselves is, however, beyond the scope of this thesis.

ppGpp

While sigma factors compete directly with each other for binding to the core RNA polymerase, ppGpp is believed to indirectly affect the competitive ability of many sigma factors (Jishage et al. 2002; Laurie et al. 2003; Magnusson et al. 2005; Dalebroux & Swanson 2012). ppGpp is a nucleotide-based secondary messenger, and the effector molecule of the stringent response. The stringent response is the transcriptional response to limiting nutrients, which ensures a rapid reallocation of cellular resources by stopping the synthesis components used for growth and cell proliferation, and in turn activating the production of factors that are crucial for stress resistance and amino acid synthesis (Durfee et al. 2008; Dalebroux & Swanson 2012).

Under non-stressful conditions, ppGpp is maintained at minimal levels in the cells. However, when facing nutrient starvation, ppGpp is produced from the ribosome associated ppGpp synthases, RelA and SpoT, to start producing ppGpp (Haseltine & Block 1973). ppGpp exerts its action through a number of different mechanisms, many of which specific molecular details still need to be unravelled. A general consensus is, however, that it indirectly modulates sigma factor competition and redirects transcription from the majority of genes controlled by RpoD involved in growth, and instead directs transcription to genes controlled by alternative sigma factors involved in amino acid synthesis and stress responses (Österberg et al. 2011). Some evidence points to the fact, that at least one of the outcomes of ppGpp is to lower the affinity for RpoD to the core, thus remodelling sigma factor competition and allowing alternative sigmas to compete for the core RNAP machinery (Hernandez & Cashel 1995; Hernandez & Cashel 1995; Cashel et al. 2003).

Several *in vitro* studies also point to ppGpp as having a sigma factor competition modifying function. For example, it has been shown that ppGpp *in vitro* reduces the ability of RpoD to compete for core RNAP against RpoH, the sigma factor required for the normal expression of heat shock genes, that *in vivo*, RpoH and RpoS display a decreased affinity for core RNAP compared to RpoD in the absence of ppGpp (Jishage et al. 2002), and that ppGpp is involved in regulation of RpoN transcription (Laurie et al. 2003; Bernardo et al. 2009), though not necessarily directly. Direct involvement of ppGpp in concert with its potentiator DksA has also been suggested for AlgT mediated transcription (Costanzo & Ades 2006; Costanzo et al. 2008).

A model that explains the function of ppGpp as altering sigma factor competition directly is however, difficult to unite with recent high resolution crystal structures showing that ppGpp

binding to the *E. coli* RNA polymerase does not involve RpoD. Rather, ppGpp binds at the interface between the β' and ω subunits (Ross et al. 2013; Zuo et al. 2013).

Direct remodelling of sigma factor competition and a lowered affinity of RpoD to the core RNAP is further in contrast to the fact that in situations of high ppGpp concentrations, genes involved in amino acid synthesis are upregulated, while genes involved in ribosomal RNA production are inhibited (Paul et al. 2005). As both classes of genes are controlled by RpoD, the differential regulation of these genes due to ppGpp presence indicates that ppGpp not necessarily modulates RpoD structure, or its affinity to core RNAP, but rather exert a promoter specific effect. This promoter specific effect is suggested to be dependent on the presence of a GC-rich discriminator in the region between the -10 box and the transcription start site, as well as the length of the linker between the -35/-10 region (Potrykus & Cashel 2008).

It is by now clear that ppGpp is an important signalling molecule, best known for its role in the bacterial stress response. However, much of its direct function still needs unravelling. Regardless of the mechanism of action exerted by ppGpp, the fraction of RpoD bound to core RNAP *in vivo* is lower when ppGpp is present, compared to when ppGpp is absent (Hernandez & Cashel 1995), meaning that, whether directly or indirectly, ppGpp may somehow influences sigma factor competition and global gene expression.

Chapter 4

Transcriptional Regulatory Networks

Recent years´ technological advances mean that studying bacterial gene regulation is not limited to studying one TF or DNA regulatory sequence at a time. With powerful techniques such as comparative genomics, transcriptomics, and high throughput mapping of *in viv*o TF-DNA interactions, researchers have been able to produce models of how-large scale TRNs are structurally organised and function *in vivo*. Whereas the previous Chapter provided insight into a number of different regulatory elements individually, the following Chapter will provide an introduction to how these different entities interact to create the complex interplay that constitute TRNs. The Chapter has been written with the realisation that this presentation is in no way exhaustive, however, the most important aspects relating to the aim of this thesis are dealt with.

A TRN describes the resulting gene expression as a function of regulatory inputs specified by interactions between proteins and DNA (Blais & Dynlacht 2005). TRNs are highly complex structures and characterised by a multilayer structure that allows for an advanced and precise cellular response to environmental factors, a response that is much more variated and complex than could be performed by the single TFs alone. Having a fine-tuned TRN is thus a perfect way of utilising the already existent capacities lying within the genetic material without introducing modifications that become fixed in the population. TRN representations are thus illustrations of network graphs that visualise molecule-molecule interactions and how the relationship between these dictates cellular behaviour (Blais & Dynlacht 2005).

A TRN is arranged in different levels (Figure 6). The first level (Figure 6A, red nodes) consists of global regulators, often sigma factors, e.g. RpoD, that act as hubs (that bind a disproportionately large number of target genes (Macneil & Walhout 2011) and control expression of a large number of genes, either directly or through a second level (Figure 6A, orange nodes), which is composed of a local network with a local regulator that controls a smaller number of genes. The third level (Figure 6A, blue nodes) consists of single genes that may interact with each other (Brooks et al. 2011).

Given their important role in bacterial adaptation and the ability to respond to continuous environmental changes, it is not surprising that TRNs are highly complex structures, and that the actual regulatory response of any signal is much more complex than the typical schematic drawings depicted in Figure 6.

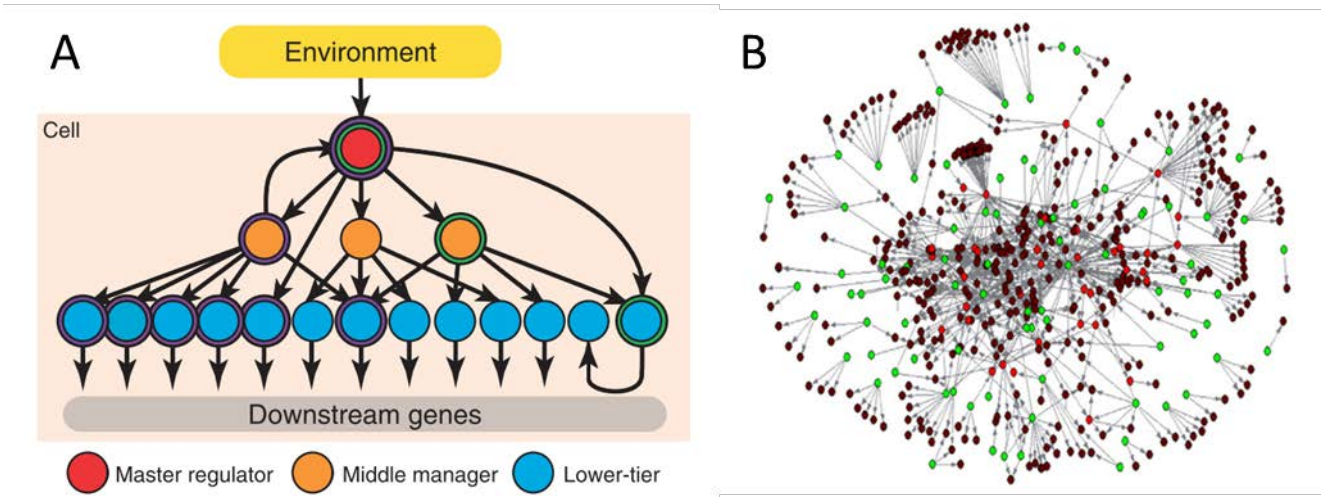


Figure 6. Structure of a TRN. (A) Schematic illustration of the organisation of a TRN displaying how, in a simplified form, a global regulator (red nodes) controls the expression of other regulators (orange nodes), which in return control the output of a number of structural genes (blue nodes), or control a number of structural genes directly. (B) Representation of the *E. coli* transcriptional regulatory network. Green circles represent TFs, brown circles denote regulated genes, and those with both functions are coloured in red (Guzman-Vargas & Santillan 2008; Brooks et al. 2011).

4.1 Modelling of Transcriptional Regulatory Networks

As system-wide data become available for an increasing number of organisms, regulators and under different experimental conditions, computational methods have been developed to integrate experimental genome-wide data into intuitive models that can be used as a starting point for further exploration. Some computational methods integrate *in vivo* generated data, while others rely solely on *in silico* analysis and comparative genomics to generate new TRN models (Brooks et al. 2011).

4.1.1 *In silico* modelling of Transcriptional Regulatory Networks

While it is beyond the scope of this thesis to provide the reader with an understanding of how *in silico* modelling is performed in detail, it is important to provide an overview of the approaches that

are used, as they each present advantages and limitations when introducing concepts such as evolution of TRN, mutations, epistasis, and pleiotropy.

Most *in silico* modellings of TRN rely on comparative genomics. The first step in these analyses is to identify orthologue TFs (homologous genes that diverged from a common ancestor) and to identify conservation of TF binding sites (TFBS), either amongst a set of known co-regulated genes, or within homologous promoters of closely related species (Price et al. 2007). While these approaches are invaluable as visualisations and starting points for further experimental analyses, it has been shown that they suffer from the limitation that orthologous TFs may sense different signals and regulate different pathways (Price et al. 2007). Other limitations to *in silico* modelling are that the effects of varying conditions such as stress conditions, other species interactions, or host-interactions cannot be predicted, just like effects created by mutations, pleiotropy, and epistasis are difficult to predict. These limitations may be addressed by experimental data, which can then be incorporated into the existing models (Karlebach & Shamir 2008).

4.1.2 Experimental based construction of transcriptional regulatory networks

Until the introduction and general use of NGS, experimental methods for investigating TF activities and mechanisms relied on low throughput *in vitro* methods such as *in vitro* transcription and electrophoretic mobility shift assay (EMSA). With microarray and NGS technology, entire genome-wide TF binding profiles can be investigated with chromatin immunoprecipitation sequencing (ChIP-seq), gene expression profiling (mRNA-seq), and high throughput *in vitro* methods such as DNA binding arrays and *in vitro* DNA immunoprecipitation-sequencing (DIP-ChIP) (Geertz & Maerkl 2010).

Especially ChIP-seq coupled with mRNA-seq has become a popular tool for experimental elucidation of TRN structures (Figure 7). For example, genome-wide binding profiles of the sigma factor σ^{54}/RpoN has been characterised in both *E. coli* (Bonocora et al. 2015), *Vibrio Cholerae* (Dong & Mekalanos 2012), and *P. aeruginosa* (Schulz et al. 2015). These studies act as important additions to *in silico* models, as they add valuable knowledge that would not have emerged from pure comparative genomics alone. For example, when studying genome-wide binding of RpoN in *E. coli*, it was suggested that RpoN could serve additional roles beside its well documented role in promoter recognition and transcription initiation (Bonocora et al. 2015). The authors found that RpoN binds an unexpected large number of intragenic regions and suggested that these regions are likely to be functional. While the exact functions of these intracellular binding regions remain to be

unravelled, findings like these are important, as traditional *in silico* modelling of TRN could have rejected them based on the fact that they are far away from a transcription start site (TSS).

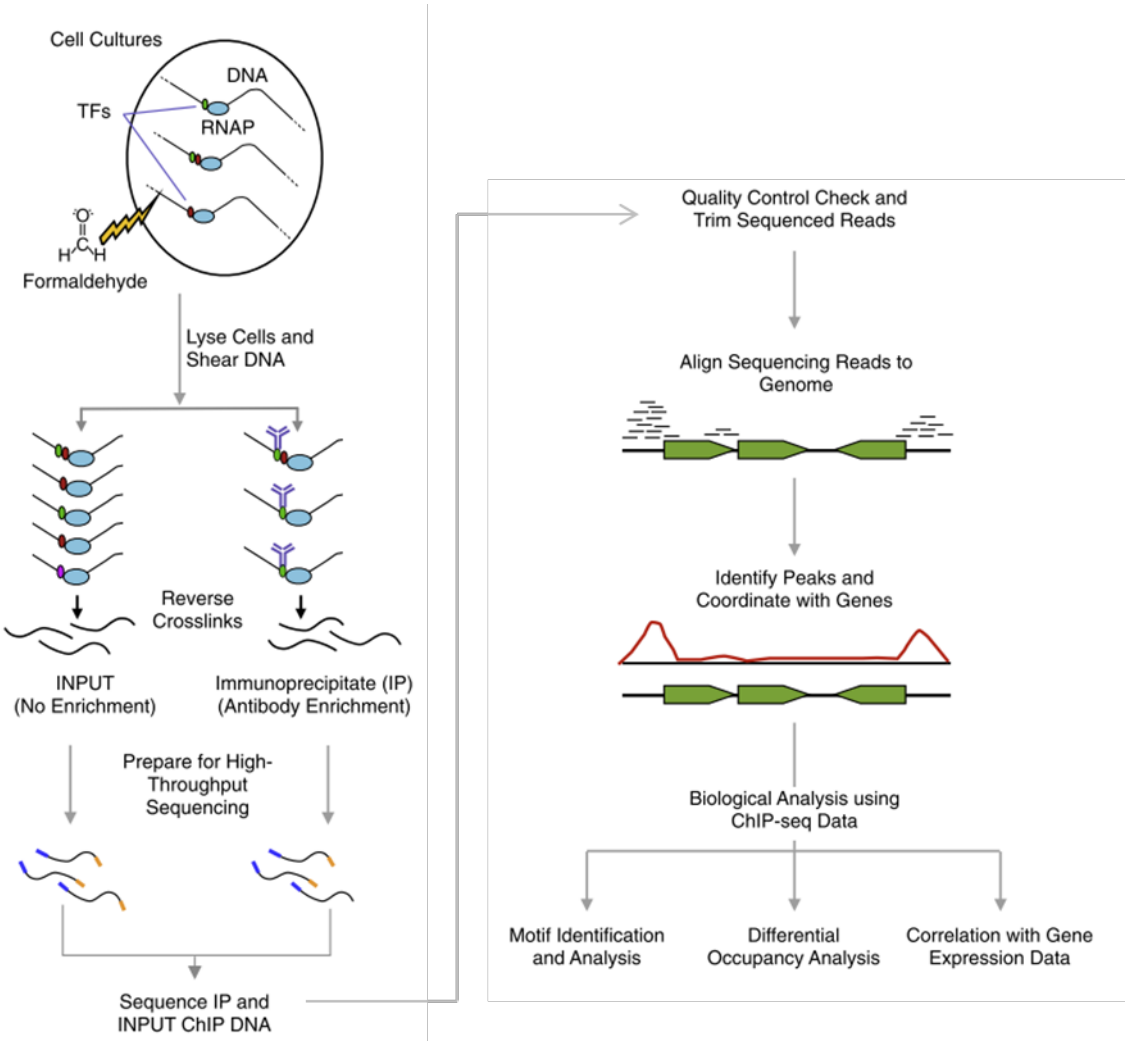


Figure 7. **Schematic overview of the steps involved in ChIP-seq**. Briefly, cells are proliferated in certain conditions until a predetermined growth phase is reached. Formaldehyde is then added to crosslink proteins to DNA. Cells are lysed, and DNA is sheared by sonication, after which DNA-TF complexes of interest are enriched using antibodies that are specific to the TF being studied. After release of the TF from DNA, and after isolation of DNA, the sample contains only the specific DNA fragments that interacted with the TF at the time of crosslinking. By sequencing the DNA, sequences can thus be mapped to the genome, and areas of the genome bound by the TF are visualised as having a higher coverage than unbound regions (Myers et al. 2015).

4.2 Evolution of Transcriptional Regulatory Networks

Understanding how TRNs are shaped and evolve is an important aspect of evolution, since the specific evolutionary effects that contribute to the shaping of TRNs (mutations, gene duplications and deletions, and HGT) extremely rarely, if ever, result in uniform effects on all genes, or are limited to one regulon. HGT results in the transfer of entire functional modules, e.g. transfer of elements that enable antibiotic resistance mechanisms in bacteria (Davies & Davies 2010). Gene duplication also results in evolution of TRN, and a subsequent functional divergence of duplicated genes may further contribute to the development of new cellular functions (Brooks et al. 2011). Mutations that shape TRN may occur in both TFs as well as in cis-regulatory elements (regions of non-coding DNA which regulate the transcription of nearby genes) and in downstream target genes (Brooks et al. 2011). Mutations in TFs may change the ability of the TF to interact with other regulators, or they may change the ability to interact with, or recognise TFBS, while mutations in cis-regulatory elements may modify, create, or delete new TF binding sites. While it may be less complicated to infer about the role of HGT, e.g. when transfer of entire genetic islands coding genes for antibiotic resistance is observed, it is more troublesome to infer about the possible roles of SNPs and single amino acid deletions in evolution of TRNs.

An example of the difficulties of predicting the effects of single amino acid mutations on entire TRN is exemplified by the transcriptional terminator Rho, in *E. coli*. While Rho dependent transcriptional termination insulates the cell from deleterious expression of prophage and other horizontally acquired DNA, as well as safeguards genomic integrity in regard to the transcriptional and replication machinery, a mutation in Rho was found to cause both an expected overexpression of a number of genes, but also a unexpected and relatively large number of genes that were down regulated, illustrating the indirect effects of the Rho mutation on TRNs. Furthermore, the authors found that the Rho mutation provided both direct fitness effects, as well as fitness effects arising from positive epistatic interactions, which underlines the potential of the Rho mutation to open evolutionary paths that would otherwise be inaccessible (Freddolino et al. 2012).

Consistent with the findings of (Freddolino et al. 2012), it is well established that especially global regulators are target for evolution, and given their important role as network regulators, any alteration of their function is expected to create large effects that may be overall beneficial, but also may have certain maladaptive side-effects, which in turn can be reduced by subsequent compensatory mutations (Hindré et al. 2012).

This process of TRN evolution, emerging maladaptive side-effects, and subsequent compensatory mutations is depicted in Figure 8, p. 27. While one mutation in one global regulator may cause a number of unexpected phenotypes, known as pleiotropy, and may complicate any phenotype prediction based on genotypic data, this task is even further complicated by epistatic effects. Epistatic effects are interactions between two or more mutations that combined produce a phenotype which deviates from the sum of the effect of the individual mutations (Elena & Lenski 2003). An example of the complexity and unpredictability of epistatic effects in evolution of TRN is *P. aeruginosa* adaptation to the CF lung environment. While it has been confirmed that global regulators are indeed also target for early adaptive mutations in this natural system (Yang, Jelsbak, Marvig, et al. 2011), a follow-up study showed that when introducing an exact combination of four of these mutations into an isogenic *P. aeruginosa* strain, an unexpected rise in antibiotic resistance was observed. This indicates that global regulator mutations not only shape their immediate TRN, but also that epistatic effects may give further rise to an increase in antibiotic resistance. A persistent and unresolved problem when studying TRNs is thus the genotype-phenotype relation and how to accurately predict phenotypic outcomes based on genotype data.

Due to its role as an opportunistic pathogen, and its role in bacterial colonisations and chronic infections in the CF lung, *P. aeruginosa* represents a specific case where evolution of TRNs in a natural system can be studied on a time scale estimated to be bacterial 200,000 generations (Yang, Jelsbak, Marvig, et al. 2011). During this time, a fluctuating environment, sub-niches, interacting species that also evolve, antibiotic treatment, and host immune defences all shape *P. aeruginosa* TRNs, which thus serve as a model case for studying natural evolution of TRNs.

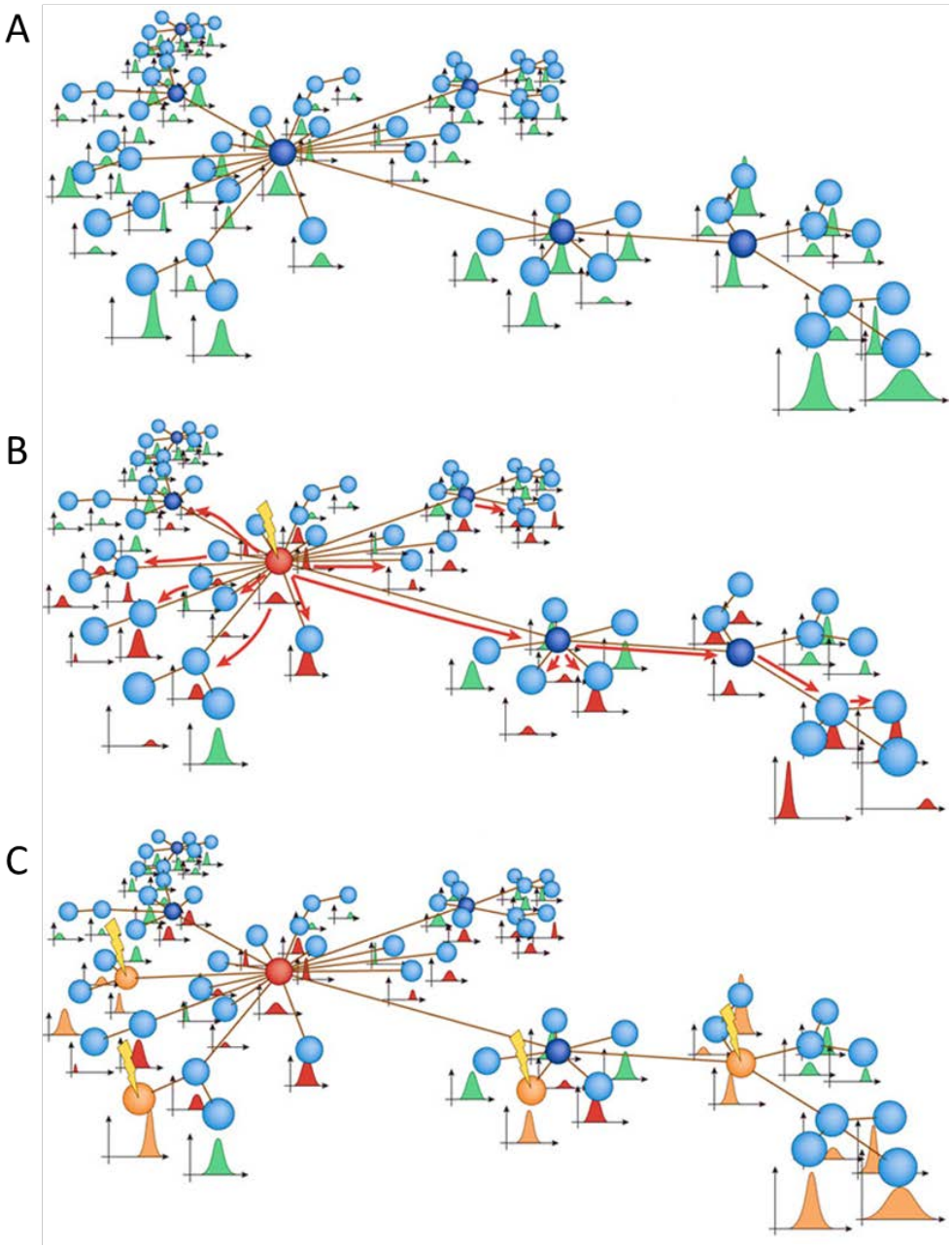


Figure 8. Evolution of a TRN. The ancestral network (A) is composed of global regulators (dark blue) and regulated genes (light blue) and the expression profiles of the regulated genes (green curves) define the phenotype. In an evolved network (B), a mutation (yellow arrow) in the global regulator alters the transcription profile of a number of genes (red arrows and curves). The overall effect of the mutation is beneficial, but negative pleiotropic effects drive subsequent compensatory mutations (orange circles) in lower-level regulatory genes, and restore expression levels to a more optimised overall signature (Hindré et al. 2012).

4.3 *Pseudomonas aeruginosa* transcriptional regulatory networks

A large number of studies have characterised TRNs in *P. aeruginosa*, both experimentally as well as by *in silico* modelling. *In silico*, the structural and functional properties of a network representing 12% of the total 5,570 genes and 16% of the predicted 500 regulatory proteins (including 54% of the 26 sigma factors) have been elucidated (Galán-Vásquez et al. 2011). This study found that the regulatory functions are biased towards particular biological processes involved in pathogenesis and virulence, such as alginate and biofilm formation, production of virulence factors, and antibiotic resistance, many of which are coordinated by quorum sensing in the bacterial population (Galán-Vásquez et al. 2011).

Experimentally, a large number of major regulatory networks have been studied in *P. aeruginosa* using both ChIP-seq, mRNA-seq, as well as with traditional low-throughput methods. The combination of ChIP-seq coupled with mRNA-seq has been used to study the key transcriptional regulator AlgR and its regulation of the sigma factor AlgT and quorum sensing (QS)-regulated virulence factors (Kong et al. 2015). Perhaps the most exhaustive transcriptional regulatory network study in *P. aeruginosa* has been that of (Schulz et al. 2015), elucidating the complex structure of 11 sigma factors regulons in the *P. aeruginosa* strain PA14. This exhaustive study revealed that sigma factor regulons constitute a highly modular network architecture with insulated functional sigma factor modules. The regulon structure also revealed a limited, but highly function-specific, cross-talk between different sigma factors. Lastly, these data present a valuable reference dataset that can be used for future extended analyses or incorporated into *in silico* analyses.

4.3.1 The transcriptional regulatory network controlling alginate production in *Pseudomonas aeruginosa*

The TRN controlling alginate production in *P. aeruginosa* is a well-known and well-studied regulatory network**.** The following section will use the TRN controlling alginate production in *P. aeruginosa* as a specific example of the complexity of TRNs, as well as an example of how CF driven adaptation may promote remodelling of TRNs through global regulator mutations.

Overproduction of alginate produces a mucoid phenotype that is protected from various host defences (Leid et al. 2005) and antibiotics (Govan & Fyfe 1978). The alginate producing phenotype marks the transition from an intermittent colonisation to chronic infection with *P. aeruginosa*, and great effort has been put into understanding the genetic regulation of alginate production, as well as the environmental factors driving the adaptation of *P. aeruginosa* towards the mucoid phenotype.

Not only is it of medical importance to achieve an understanding of the factors controlling alginate expression, it also serves as a prime example of how regulatory networks may be shaped and remodelled by evolution, and how several global regulators compete and interact to control the expression of a very defined phenotype. In addition, it serves as an example of the complexity variations of TRNs that are difficult to model *in silico*, since the collective evidence from a number of studies have shown that evolutionary shaping of this network works through a number of different mechanisms, such as sigma factor point mutations, anti-sigma inactivation, and perhaps even modulation of ppGpp levels.

The production of alginate is regulated through a number of different regulatory pathways, of which most converge to a common pathway, controlled by the alternative ECF sigma factor, AlgT (σ^{22}). Alginate production is initiated from the *algD* promoter controlling the 12 gene *algD* operon (Chitnis & Ohman 1993; Schurr et al. 1993) and is activated by stress-conditions, such as antimicrobial and oxidising agents, elevated temperatures, and osmotic imbalances. Under non-inducing conditions, AlgT is sequestered by its anti-sigma factor, MucA, and its complex-forming partner, MucB. The binding of AlgT to MucA prevents transcription from the *algD* promoter (Xie et al. 1996) (Figure 9A). In response to envelope stress, a signal transduction pathway confers proteolytic cleavage of MucA and MucB, thereby releasing AlgT and activating it for alginate production through transcription from the *algD* promoter (Qiu et al. 2007) (Figure 9B). Release of AlgT from the MucA-MucB complex is not an absolute determinant for subsequent AlgT transcription from the *algD* promoter. After release of AlgT, a number of other factors determine whether AlgT gains access to the *algD* promoter and initiates transcription (Figure 9C). For example, an RpoN promoter sequence which directly overlaps the AlgT dependent promoter may cause antagonistic effects and inhibit AlgT dependent transcription from the *algD* promoter *in vitro* (Boucher et al. 2000). Also sigma factor competition between AlgT and RpoD (which is further regulated by its anti-sigma factor, AlgQ) impacts the transcriptional activity of AlgT (Yin et al. 2013).

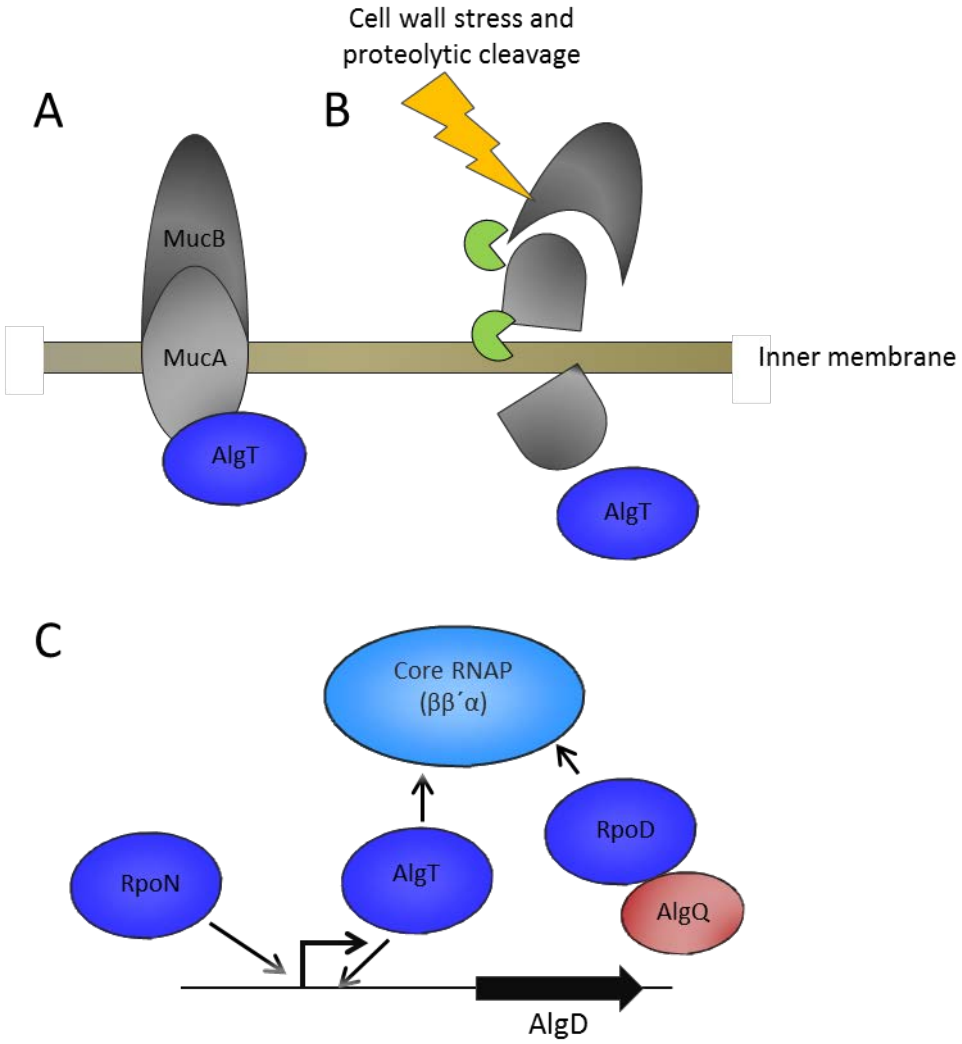


Figure 9. Genetic regulation of alginate production in *P. aeruginosa*. Expression of the *algD* operon is a highly regulated and complex process. (A) Under non-induced conditions, AlgT is sequestered by the MucA-MucB complex. However, both evolutionary rewiring and environmental factors such as cell-wall stress may cause proteolytic degradation of the MucA-MucB complex, thereby releasing AlgT (B). AlgT is free to initiate transcription from the *algD* promoter to initiate alginate synthesis. Transcription from the *algD* promoter is, however, a complex interplay between sigma factor antagonism from RpoN binding at overlapping promoter sequences, as well as sigma factor competition between AlgT and RpoD, to gain access to the core RNA polymerase.

ppGpp controlled regulation of the AlgT regulatory network

A number of studies have suggested that activation of the AlgT TRN may be controlled by a secondary pathway, independent of envelope stress mediated proteolytic release of AlgT from the

MucA-MucB complex. This cascade is initiated as a response to metabolic stress, where the accumulation of the alarmone ppGpp directly stimulates AlgT activity during entry to stationary phase (Costanzo & Ades 2006; Costanzo et al. 2008; Gopalkrishnan et al. 2014). These studies found that the effect of ppGpp and its cofactor DksA was an increased transcription form certain AlgT controlled promoters both *in vivo* and *in vitro*, but that the production of AlgT itself was not affected by ppGpp. This indicates that the promoter specific effect of ppGpp is not limited specifically to RpoD controlled promoters, as is the generally accepted model. Instead, ppGpp may present a direct way of remodelling transcription through promoter specific effects on complexes that are sensitive to ppGpp, such as certain members of the AlgT regulon.

4.3.2 Evolution of the transcriptional regulatory network controlling alginate production in *Pseudomonas aeruginosa*

During adaptation of *P. aeruginosa* to the CF lung environment, the AlgT regulatory network is continuously remodelled. The mucoid phenotype is a hall mark of chronic infection (Koch & Høiby 1993), and both mucoid and non-mucoid phenotypes of *P. aeruginosa* are frequently isolated simultaneously from CF lungs (Thomassen et al. 1979; Shawar et al. 1999). Overproduction of alginate is highly demanding for the cell, and the cycling between mucoid and nonmucoid clinical isolates indicates that alginate production may produce negative side-effects that require continuous remodelling of the regulatory network, much in line with the description of the generalised "TRN evolution model" presented in Figure 8, page 27. While alginate production in an environmental strain is initiated by proteolytic degradation of MucA, appearance of mucoid *P. aeruginosa* clinical isolates is most often a result of mutations in MucA that inactivate the protein, thereby releasing AlgT from the MucA-MucB complex (Martin et al. 1993; Folkesson et al. 2012).

A subsequent reversion to a non-mucoid phenotype is mediated by a number of different mutations in the AlgT gene (Jelsbak et al. 2007; Yang, Jelsbak, Marvig, et al. 2011; Damkiaer et al. 2013), and it is speculated that mutations in the core binding domain of AlgT may affect the binding affinity to the core RNAP, thereby affecting the ability of AlgT to compete against RpoD for core RNAP. Involvement of sigma factor competition between AlgT and RpoD was also suggested by (Yin et al. 2013), who showed that overexpression of the RpoD anti-sigma factor AlgQ gave rise to a mucoid phenotype due to sequestering of RpoD by AlgQ, thereby enabling AlgT to form holo RNAP and activate alginate production.

Interestingly, the evolutionary tinkering of the AlgT regulatory network does not end with the non-mucoid phenotype of a MucA-AlgT mutant. Subsequent mutations in the housekeeping sigma

factor RpoD has been shown to cause a direct transition to a mucoid phenotype (Damkiaer et al. 2013). The RpoD mutation was isolated from *P. aeruginosa* clinical CF isolates a significant period of time after emergence of the AlgT mutation, and thus illustrates that evolution of TRNs is an ongoing process that constantly reshapes and optimises the bacterial TRN to respond to new challenges. By introducing each of the *mucA*, *algT*, and *rpoD* mutations in a PAO1 strain, the authors were able to show that these precise mutations were the direct cause of the phenotype changes (Figure 10). As a number of different mutations in the genes encoding *mucA* and *algT* may cause mucoid/nonmucoid phenotype changes (Ciofu et al. 2008), the specific regulatory mutations discovered in the CF related *P. aeruginosa* DK2 strain are denoted with a "DK2" superscript.

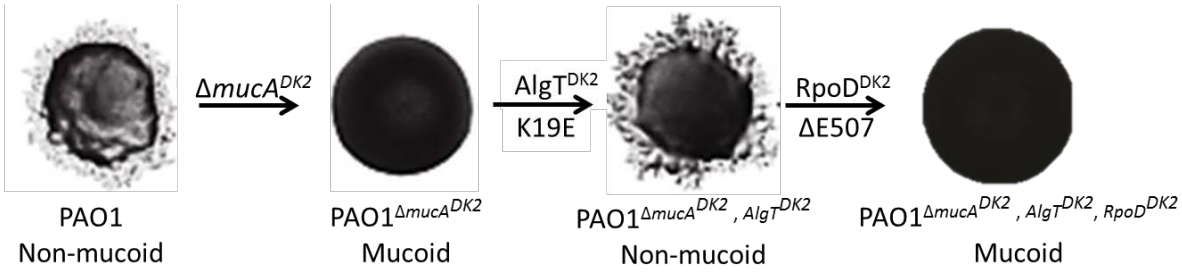


Figure 10. **Switch in mucoid/nonmucoid phenotype through remodelling of the AlgT TRN**. Initially, a non-mucoid *P. aeruginosa* PAO1 is non-mucoid, but in the CF lung environment, a mutation causes inactivation of MucA, which causes a switch to a mucoid phenotype. Subsequently, a mutation in the core RNAP binding region of AlgT causes a shift to a non-mucoid phenotype, and after a period of time, appearance of mucoid clinical isolates with mutations in the housekeeping sigma factor have been isolated. Modified from (Damkiaer et al. 2013).

The above presented evolutionary path that enables *P. aeruginosa* to shuffle between the mucoid and non-mucoid state is centralised on the initial mutation causing MucA inactivation. While this is the most common path leading to mucoidy in *P. aeruginosa* CF clinical isolates, few examples have presented evidence that inactivation of MucA is not necessarily a prerequisite for alginate overproduction. Mutations in the AlgT negative regulators MucB and MucD discovered in *P. aeruginosa* clinical CF isolates have also been shown to cause moderate alginate production (Ciofu et al. 2008), as well as a mutation found in the sensor kinase KinB has been shown to cause a RpoN dependent alginate overproduction of alginate (Damron et al. 2012). Common for those are that

even though MucA is not directly inactivated, they result in proteolytic degradation of MucA, thereby releasing AlgT for subsequent alginate production.

With the TRN controlling alginate production in *P. aeruginosa* as an example, it is obvious that a great deal of knowledge has already been gathered on the function and organisation of TRN. However, common for the major part of these studies is that the specific molecular mechanisms connecting the evolutionary changes to the resulting phenotypic are unknown. Therefore, to achieve a complete understanding of how mutations may remodel regulatory networks during adaptation to novel niches, we must be able to explain the molecular effects of these mutations.

Chapter 5

Present Investigations

In chapter 1, the reader was introduced to bacterial evolution, and how deletions, acquisitions and point mutations in the genetic material comprise the raw material of evolution that enables bacteria to evolve and adapt, and facilitate colonization of novel environments. The reader was introduced to the concept of evolution experiments as a mean to study bacterial evolution, and how certain natural systems provide the opportunity to study bacterial evolution in its natural context.

Chapter 2 gave insight into areas of bacterial gene regulation with a focus on transcription and the factors involved in its regulation, such as sigma factors, anti-sigma factors, sigma factor competition and ppGpp involvement in transcriptional control. Special attention in this chapter has been placed on sigma factors and their regulation, since preceding bioinformatics analyses from this group have shown that especially sigma factors are target for early adaptive mutations in *P. aeruginosa*.

In chapter 3, the concepts of evolution and transcription were combined to provide the reader with an introduction of TRNs and evolution of TRNs. Important terms such as pleiotropy and epistasis were introduced, and special attention was payed to presentation of the *P. aeruginosa* TRNs, exemplified by the regulatory network controlling alginate production controlled by a global regulator, the sigma factor AlgT.

5.1 Background

The research in this thesis rests heavily on previous bioinformatics analyses of the Copenhagen *P. aeruginosa* collection of clinical isolates, sampled from CF patients since 1972. Intensive bioinformatics analyses have gone into characterising these isolates, which have resulted in a characterisation of two strains that have spread among the CF patients; DK1 and DK2, respectively. A number of bioinformatics analysis have characterised the evolution of especially the DK2 strain in respect to rates and detection of SNP mutations (Yang, Jelsbak, Marvig, et al. 2011), changes in the genetic composition during evolution (Rau et al. 2012), identification of pathoadaptive genes (Marvig et al. 2013) as well as the phenotypic impact of mutations fixed in global regulator genes

was illustrated by the reconstruction of just four global regulator mutations that showed striking phenotypic similarities to the evolved DK2 clone (Damkiaer et al. 2013). These studies not only highlight the evolvability of the *P. aeruginosa* DK2 clone, they also confirm that especially global regulators are targets for adaptive mutations, that these mutations result in extensive remodelling of TRNs which produce epistatic effects and phenotypes.

5.2 Aim of thesis

Given the in-depth knowledge of the DK2 strain, its evolutionary history and knowledge about the mutations in certain important regulators, but also the lack of knowledge about the specific molecular mechanisms of these mutations, the focus of this thesis has been to study the molecular effect of mutations in 3 global regulators, as well as the overall impact on their corresponding TRNs.

The specific aim of this thesis is therefore

- **To determine the specific molecular mechanisms of the mutations in the global regulators AlgT, RpoD and RpoN.**
- **To determine how the altered mechanisms of these global regulators affect their entire TRN.**

5.3 Outline of studies

Paper 1: In the first paper, we examined the molecular consequences, as well as the impacts on global TRNs, arising from a single amino acid substitution in the σ^{54} sigma factor RpoN, discovered in the *P. aeruginosa* DK2 lineage. We used a combination of ChIP-seq, mRNA-seq as well as *in vitro* protein-DNA interaction studies, to systematically investigate the functional consequences of the amino acid substitution in $RpoN^{DK2}$, and we demonstrated that the direct effect of the mutation was a decreased ability of $RpoN^{DK2}$ to bind to its promoter sequences. The effect was translated into a loss of regulatory connections and a downregulation of the majority of genes controlled by RpoN.

Interestingly, the mutation was not comparable to a knock out variant. The evolved $RpoN^{DK2}$ network gained a regulatory capacity of the *tad* (tight adherence) locus, which is involved in biofilm formation, colonisation, and pathogenesis in a wide range of bacterial species (Tomich et al. 2007). RpoN is not previously been assigned any functional control over this specific region, and this result thus reflect the complex evolutionary rewiring of regulatory networks. In addition the added regulatory capacity of the $RpoN^{DK2}$, we observed an enhanced crosstalk effects to other sigma

factor regulons, confirming that mutations in regulators produces complex effects on neighboring networks.

The work presented in Paper 1 provides a molecular explanation of a naturally occurring global regulator mutation, but it also demonstrates how these specific effects can be translated into a remodeling of several global TRNs.

Paper 2: In paper 2, we set out to investigate the underlying molecular mechanisms of evolutionary modifications of the TRN controlling alginate production in *P. aeruginosa*. As mentioned, this regulatory network has been extensively shaped by global regulator mutations, and the regulatory mechanisms involved in controlling alginate production are well-studied on a genotype-phenotype level, but poorly understood on a molecular level. It thus presents an excellent opportunity to study in detail, how mutations in several global regulators each and in combination shape and rewire their respective networks. While the regulatory mechanisms controlling alginate production extends further than AlgT and RpoD, we specifically chose these two based on observations from (Damkiaer et al. 2013). This study showed that an evolutionary induced shuffling between a nonmucoid and mucoid phenotype was dependent consecutive mutations in the ECF sigma factor AlgT and housekeeping sigma factor RpoD (Illustrated in Figure 10, p. 32).

We therefore set up a systematic study in which the molecular consequences of each regulator was studied *in vitro* by SPR, and *in vivo* by gene expression profiling, ChIP-seq and by regulation of sigma factor levels. By studying the regulatory response of each mutation, as well as the combination of mutations, i.e. the PAO1$^{\Delta mucA,\ algT(DK2),\ rpoD(DK2)}$ strain, we were able to assign regulator specific alterations, as well discover which regulatory modifications arose from epistatic effects.

Our results confirm that the combinatorial effect of mutations in global regulators results in complex epistatic effects and that mutations in AlgT and RpoD cause very different changes in the proteins functions, although both protein mutations are required to produce the evolved mucoid phenotype. The AlgT^{DK2} mediated switch from the mucoid to nonmucoid phenotype resulted from a direct decrease in the affinity between the AlgT^{DK2} and the core RNAP, measured by SPR, and this decreased affinity caused the AlgT^{DK2} protein to compete less efficiently for the core RNAP, thus abrogate transcription from the *algD* promoter, as well as cause a consistent downregulation of genes in the AlgT regulon.

Further remodelling caused by the $RpoD^{DK2}$ protein was found to involve complex epistatic effects. Interestingly, the $RpoD^{DK2}$ protein was unaffected in both its ability to interact to the core RNAP and to recognise RpoD controlled promoter sequences. We also observed only subtle effects on the gene expression analysis when we compared gene expression values of $PAO1^{rpoD(DK2)}$ to $PAO1^{rpoD(WT)}$. Regulatory effects from the $RpoD^{DK2}$ appeared only in combination with the *ΔmucA, algT(DK2)* mutations. We thus concluded that the effects produced by the $RpoD^{DK2}$ protein were contingent on the genetic environment and the presence of the $PAO1^{\Delta mucA,\ algT(DK2)}$ combination, and could involve an unknown factor not recognised by the gene expression or ChIP-seq data. Based on sequence alignment of *P. aeruginosa* RpoD and *E. coli* RpoD, we suggested that this unknown factor could be an altered sensitivity to the signal molecule, ppGpp, as the $RpoD^{DK2}$ mutation is found in close proximity to two *E. coli* RpoD variants, discovered as suppressor alleles of auxotrophy in a $ppGpp^0$ strain (Hernandez & Cashel 1995).

We thus propose a model in which initial rewiring of the AlgT regulatory network is mediated by a decrease in affinity between the $AlgT^{DK2}$ and core RNAP. The result is a direct downregulation of genes controlled by AlgT, including the entire alginate biosynthesis operon. The subsequent remodelling of several regulatory network, is caused by an increased or permanent sensitivity of $RpoD^{DK2}$ to an unknown factor, presumably ppGpp, and this epistatic effect causes an indirect remodelling of sigma factor competition, which gives rise to the mucoid phenotype of $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$.

Paper 2 thus links the direct, functional effects of two mutations in global regulators to the resulting network displacement, epistatic effects and resulting phenotype, and the results from this paper demonstrate the need for achieving a deeper understanding of the evolution of these complex networks.

Chapter 6

Conclusion and future perspectives

During the past years, studying the evolution of bacterial TRNs has greatly benefitted from the development of high throughput experimental techniques, as well as increasing computing power and sequencing techniques. The advances have allowed detailed construction of TRNs from a number of organisms. With detailed network models, focus has shifted from assigning function to individual genes, to understanding how entire sets of genes and molecules work in a complex network. A complete understanding of TRNs, however, must encompass both an understanding of specific molecular processes as well as how these individual molecular processes are correlated to form an entire network.

The aim of this was to determine the specific molecular mechanisms of adaptive mutations in the global regulators AlgT, RpoD and RpoN, and to determine how the altered mechanisms of these global regulators affect entire TRNs. By studying these 3 different transcriptional regulators, we were able to determine three specific molecular mechanisms which mediate rewiring of regulatory networks.

Investigations of the specific molecular consequences of the three sigma factor mutations $RpoN^{DK2}$, $AlgT^{DK2}$, and $RpoD^{DK2}$ revealed that sigma factor function is shaped by evolution through several mechanisms. We found that the sigma factor AlgT was modulating by a decreased affinity for the core RNAP, that the sigma factor RpoN was modulated by a decreased ability to bind its promoter recognition DNA sequences, although it gained the ability to recognise others. RpoD was remodelled by a complex mechanism for which the exact details remain to be elucidated; however, we speculate that the consequence of the mutation is an altered sensitivity to the signal molecule, ppGpp.

The evolved $AlgT^{DK2}$ sigma factor displayed a decreased affinity for the core RNAP, resulting in a specific downregulation of the entire AlgT controlled regulon, including the genes involved in alginate biosynthesis. This alteration in core RNAP binding affinity affected the entire regulon and therefore produced a uniform effect on its entire TRN.

The RpoN^{DK2} mutation affected the proteins ability to bind DNA promoter recognition sequences. Different from the AlgT^{DK2} mutation, however, this effect was not observed as a uniform response on its TRN. The mutation rather acted as a function modulating mutation, which allowed RpoN^{DK2} to bind to a so far unknown RpoN recognition sequence and direct transcription of the entire *tad*-locus. Thus, while neither RpoN^{DK2} nor AlgT^{DK2} resembles loss-of-function proteins, each of the evolved proteins display a different form of modulation, confirming the complexity of these global regulators, as well as the difficulties in predicting mutational impacts.

The RpoD^{DK2} exemplifies the challenges of studying global regulators and TRNs. The RpoD^{DK2} protein resembled the RpoD^{WT} in most of our analyses, however, it was clear that the mutation was somehow involved in producing epistatic effects when combined with the AlgT mutation, and that the mediator of these epistatic effects most likely is ppGpp. The subtle effects observed from the RpoD^{DK2} mutation could be a consequence of the nature of the protein. As a housekeeping sigma factor, the cell cannot afford to introduce permanent, drastic changes to its function. However, in combination with other global regulator mutations, small scale alteration of its function may present a way of rewire TRNs without creating detrimental side-effects.

The overall observations that can be drawn from the two studies presented in this thesis is that none of the proteins resembled biologically inactive proteins, each of the evolved molecular mechanisms differed, and each evolved mechanism resulted in a different output on their respective TRN. The findings and conclusion presented here thus not only provide important insight to evolution of specific *P. aeruginosa* TRNs. It exemplifies the difficulties of predicting evolutionary impact on entire TRN structures, as well as underlines the difficulties associated with relating genotypic changes to phenotypic outcomes.

Suggestions on future experiments

While the specific molecular mechanisms of the RpoD^{DK2} mutation still remain unclear, the results presented have paved the way for a set of experiments which could clarify the possible role of ppGpp in the TRN rewiring. Any additional experiments should focus in elucidating the possible role of ppGpp, and whether RpoD^{DK2} mimics the presence of ppGpp, whether it becomes oversensitive to ppGpp, or if another ppGpp mediated mechanism is involved.

Two *E. coli* RpoD variants were found to be suppressor alleles of a ppGpp^0 auxotroph phenotype. A straightforward method for testing the relation between these two mutations, and the RpoD^{DK2} would be to test if the RpoD^{DK2} mutation was able to supress auxotrophy, as was the case for the

two RpoD suppressor alleles in *E. coli*. However, in a $ppGpp^0$ background, *P. aeruginosa* does not display auxotrophy for any amino acids (data not shown). While this is surprising, and highlights the relevance of species-specific differences in studying TRNs, a similar situation has been discovered in *Pseudomonas putida*, which also remains prototrophic even in a $ppGpp^0$ background (Bernardo et al. 2009)

A number of alternative experiments could identify any ppGpp mediated role in the regulatory function of $RpoD^{DK2}$. By measuring ppGpp levels using the method of (Bergman et al. 2014), it could be identified if direct altered levels of ppGpp were involved in transcriptional regulation. Alternatively, comparing *in vitro* transcription profiles of the $RpoD^{WT}$-RNAP and $RpoD^{DK2}$-RNAP at different promoters (rRNA, amino acid synthesis and the *algD* promoter) could identify if the $RpoD^{DK2}$ mutation mimicked the presence of ppGpp, and altered the transcription properties of the $RpoD^{DK2}$-RNAP complex. Alternatively, if the $RpoD^{DK2}$ is indeed dependent on, or mimicking the presence of ppGpp, any changes in the regulatory properties would be expected to be more pronounces in minimal media, as our results (data not shown), as well as previous results have shown that any changes in the growth and metabolic capacity of RpoD mutants are more pronounced in minimal media.

6.3 Future perspectives

6.3.1 Transferability of results to the *Pseudomonas aeruginosa* DK2 lineage

The investigation of AlgT, RpoD, and RpoN presented in this thesis is based on experiments performed in *P. aeruginosa* PAO1. The reported results thus directly reflect the specific function of each global regulator, and are not influenced by differing experimental conditions (as would be expected in the CF lung environment) or other adaptive mutations that have accumulated in the *P. aeruginosa* DK2 lineage.

Especially this last point is of interest when discussing the future perspectives related to the results presented in this thesis. The evolutionary shaping of the DK2 lineage has produced a bacterium that is highly optimised for the CF lung environment. While the three global regulators studied in this thesis certainly demonstrate that it takes just few mutations to completely alter large TRNs of a bacterium, they represent only a subset of the actual genetic modifications accumulated in the DK2 lineage since it entered the CF lung environment.

The genome of the DK2 lineage shares 92.5 % homology with the PAO1 genome, and in addition, it contains 216 kbp of genetic material distributed on 195 genes not present in the PAO1 genome (Rau et al. 2012). During adaptation to the CF environment, the DK2 lineage experiences a total of 12 non-synonymous mutations, fixed in global regulators, as well as a total of 180 SNPs that separate the early 1973 DK2 clone to the most recent investigated DK2 clone (Yang, Jelsbak, Marvig, et al. 2011; Damkiaer et al. 2013). It is therefore likely that any of these mutations, deletions, and acquisitions will play a major part in further shaping the TRNs of *P. aeruginosa* DK2 beyond what has already been described in this thesis. This speculation therefore raises two questions;

1) Do the additional genomic features of the *P. aeruginosa* DK2 lineage result in a further and more extensive remodelling of the same TRN as studied in this thesis?
2) To which extent does species, strain, or even lineage specific TRN differentiations take place?

It would be of interest to investigate the TRN dynamics of the same global regulators as used in this thesis but perform any investigation on e.g. a laboratory strain (as used in these studies), an early adapted DK2 strain, a late adapted DK2 strain as well as a non-CF-related *P. aeruginosa*, e.g. PA14.

The analysis should subsequently be extended to include most of the global regulators known to be targeted for adaptive mutations in the *P. aeruginosa* DK2 lineage. An integration of these data into an entire TRN network model would present a first-time study of the entire evolutionary history of TRN from an actual natural and medical system, and would provide valuable knowledge of the specific impact each step of evolution has on TRNs, as well as genotype-phenotypes relationships.

6.3.2 Understanding and manipulating regulatory networks to control CF infections and fighting antibiotic resistance

With an in-depth understanding of the structure and organisation of TRNs during each step in evolution, important issues can be addressed, such as how to genetically modify organisms to optimise industrial production, or how to combat antibiotic resistance.

In 2013, the U.S. Centre of Disease Control declared the human race for “now being in the post-antibiotic era” (Centers for Disease Control and Prevention 2013), and in its 2014 report on global surveillance of antimicrobial resistance, The World Health Organization noted that “the world is

heading towards a post-antibiotic era, in which common infections and minor injuries, which have been treatable for decades, can once again kill" (World Health Organisation 2015). These challenges create a unmet need for new antibiotics, and the problem is only accelerating due to the low number of new antibiotic discoveries (Hamad 2010).

Recently, the concept of combination therapy, i.e. combining two or several drugs in one treatment, has gained attention in fighting bacterial infections. Combination therapy can be advantageous as it may produce synergistic effects (when the effect of combining two or more drugs is greater than the sum of their individual effects) as well as result in collateral sensitivity (when mutations conferring resistance to one drug increase the sensitivity to another drug) (Munck et al. 2014). Combination therapies may, however, also produce unfavourable side effects such as collateral resistance development (when a mutation conferring resistance to one drug results in an increase in resistance to another drug) (Munck et al. 2014). Unpredicted, but favourable side-effects are, of courses desirable to understand in order to reproduce the effect in other systems or in other organisms, and likewise, unpredictable, but unfavourable side-effects are, of course, desirable to avoid.

In this perspective, knowing the exact structure and function of TRNs, how TRNs may differ within closely related species, or how they may be shaped in changing environments (such as created by antibiotic pressure) is thus a true case of "knowing your enemy", the enemies taking the form of multiple drug resistant pathogenic bacteria.

The CF model system represents a perfect opportunity to combine the study of evolution in a natural system, studying the remodelling of TRNs, and the resulting development of new phenotypes, such as antibiotic resistance. Well established TRN network models could then serve as regulatory blueprints used for further studying and mapping gene expression routes leading to antibiotic resistance. With a detailed map of these routes, identifying alternative possibilities for intervention with combinatorial treatment, adjuvant strategies (i.e. drugs that increase the efficacy or potency of other drugs, but with no pharmacological effects by themselves [Pieren & Tigges 2012]) or preventing the rise of multidrug resistance by choosing the right combination for the right TRN structure could be feasible. However, before manipulations of TRNs can be used in any form for treatment strategies, there is a need for a thorough understanding of the TRN templates that might be accessible for manipulation.

Bibliography

Barrick, J.E. et al., 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268), pp.1243–1247.

Barrick, J.E. & Lenski, R.E., 2013. Genome dynamics during experimental evolution. *Nature reviews. Genetics*, 14(12), pp.827–839.

Bergman, J.M., Hammarlöf, D.L. & Hughes, D., 2014. Reducing ppGpp level rescues an extreme growth defect caused by mutant EF-Tu. *PLoS ONE*, 9(2).

Bernardo, L.M.D. et al., 2009. Sigma54-promoter discrimination and regulation by ppGpp and DksA. *The Journal of biological chemistry*, 284(2), pp.828–38.

Blais, A. & Dynlacht, B.D., 2005. Constructing transcriptional regulatory networks. *Genes & development*, (212), pp.1499–1511.

Blount, Z.D. et al., 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489(7417), pp.513–518.

Bonocora, R.P. et al., 2015. Genome-Scale Mapping of *Escherichia coli* σ54 Reveals Widespread, Conserved Intragenic Binding. *PLOS Genetics*, 11(10).

Boucher, J.C., Schurr, M.J. & Deretic, V., 2000. Dual regulation of mucoidy in *Pseudomonas aeruginosa* and sigma factor antagonism. *Molecular Microbiology*, 36(2), pp.341–351.

Brooks, A.N. et al., 2011. Adaptation of cells to new environments. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 3(5), pp.544–561.

Bryant, J., Chewapreecha, C. & Bentley, S.D., 2012. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiology*, 7(11), pp.1283–1296.

Buck, M. et al., 2000. The Bacterial Enhancer-Dependent ς54(ςN) Transcription Factor. *Journal of bacteriology*, 54(15), pp.4129–4136.

Buck, M. & Cannon, W., 1992. Specific Binding of the transcription factor sigma-54 to promoter DNA. *Nature*, 358.

Burgess, R.R. & Anthony, L., 2001. How sigma docks to RNA polymerase and what sigma does. *Current opinion in microbiology*, 4(2), pp.126–31.

Burgess, R.R. & Travers, A.A., 1969. Factor Stimulating Transcription by RNA Polymerase. *Nature*, 221.

Bye, M.R., Ewig, J.M. & Quittel, L.M., 1994. Cystic fibrosis. *Lung*, 172(5), pp.251–270.

Cashel, M., Hsu, L.M. & Hernandez, V.J., 2003. Changes in conserved region 3 of *Escherichia coli* σ70 reduce abortive transcription and enhance promoter escape. *Journal of Biological Chemistry*, 278(8), pp.5539–5547.

Centers for Disease Control and Prevention, 2013. *Antibiotic Resistance Threats in the United States*,

Cezairliyan, B.O. & Sauer, R.T., 2009. Control of *P. aeruginosa* AlgW protease cleavage of MucA by peptide signals and MucB. *Molecular Microbiology*, 72(2), pp.368–379.

Chitnis, C.E. & Ohman, D.E., 1993. Genetic analysis of the alginate biosynthetic gene cluster of *Pseudomonas aeruginosa* shows evidence of an operonic structure. *Molecular Microbiology*, pp.583–593.

Ciofu, O. et al., 2001. Characterization of paired mucoid/non-mucoid *Pseudomonas aeruginosa* isolates from Danish cystic fibrosis patients: antibiotic resistance, beta-lactamase activity and RiboPrinting. *The Journal of antimicrobial chemotherapy*, 48(3), pp.391–6.

Ciofu, O. et al., 2008. Investigation of the algT operon sequence in mucoid and non-mucoid *Pseudomonas aeruginosa* isolates from 115 Scandinavian patients with cystic fibrosis and in 88 in vitro non-mucoid revertants. *Microbiology (Reading, England)*, 154(Pt 1), pp.103–13.

Costanzo, A. et al., 2008. ppGpp and DksA likely regulate the activity of the extracytoplasmic stress factor SigmaE in *Escherichia coli* by both direct and indirect mechanisms. *Molecular Microbiology*, 67(3), pp.619–632.

Costanzo, A. & Ades, S.E., 2006. Growth phase-dependent regulation of the extracytoplasmic stress factor, SigmaE, by guanosine 3´,5´-bispyrophosphate (ppGpp). *Journal of Bacteriology*, 188(13), pp.4627–4634.

Cutting, G.R., 2014. Cystic fibrosis genetics : from molecular understanding to clinical application. *Nature Publishing Group*, 16(1), pp.45–56.

Dalebroux, Z.D. & Swanson, M.S., 2012. ppGpp: magic beyond RNA polymerase. *Nature Reviews Microbiology*, 10(3), pp.203–212.

Damkiaer, S. et al., 2013. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proceedings of the National Academy of Sciences*, 110(19), pp.7766–7771.

Damron, F.H. et al., 2012. Analysis of the *Pseudomonas aeruginosa* regulon controlled by the sensor kinase KinB and sigma factor RpoN. *Journal of bacteriology*, 194(6), pp.1317–30.

Damron, F.H. & Goldberg, J.B., 2012. Proteolytic regulation of alginate overproduction in *Pseudomonas aeruginosa*. *Molecular Microbiology*, 84(4), pp.595–607.

Davies, J. & Davies, D., 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, 74(3), pp.417–433.

Deretic, V. & Konyecsni, W.M., 1989. Control of mucoidy in *Pseudomonas aeruginosa*: Transcriptional regulation of algR and identification of the second regulatory gene, algQ. *Journal of Bacteriology*, 171(7), pp.3680–3688.

Dixon, R. & Kahn, D., 2004. Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology*, 2(8), pp.621–631.

Dong, T.G. & Mekalanos, J.J., 2012. Characterization of the RpoN regulon reveals differential regulation of T6SS and new flagellar operons in *Vibrio cholerae* O37 strain V52. *Nucleic acids research*, 40(16), pp.7766–75.

Dove, S.L. & Hochschild, A., 2001. Bacterial two-hybrid analysis of interactions between region 4 of the Sigma 70 subunit of RNA polymerase and the transcriptional regulators Rsd from *Escherichia coli* and AlgQ from *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 183(21), pp.6413–6421.

Durfee, T. et al., 2008. Transcription profiling of the stringent response in *Escherichia coli*. *Journal of Bacteriology*, 190(3), pp.1084–1096.

Elena, S.F. & Lenski, R.E., 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature reviews. Genetics*, 4(6), pp.457–69.

Farewell, A., Kvint, K. & Nyström, T., 1998. Negative regulation by RpoS: A case of sigma factor competition. *Molecular Microbiology*, 29(4), pp.1039–1051.

Ferguson, a L. et al., 2000. Interaction of sigma 70 with *Escherichia coli* RNA polymerase core enzyme studied by surface plasmon resonance. *FEBS letters*, 481(3), pp.281–4.

Flores, A.R. et al., 2015. A Single Amino Acid Replacement in the Sensor Kinase LiaS Contributes to a Carrier Phenotype in Group A Streptococcus. *Infection and Immunity*, 83, pp.4237–4246.

Folkesson, A. et al., 2012. Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nature Reviews Microbiology*, 10(12), pp.841–51.

Freddolino, P.L., Goodarzi, H. & Tavazoie, S., 2012. Fitness landscape transformation through a single Amino acid change in the rho terminator. *PLoS Genetics*, 8(5).

Galán-Vásquez, E., Luna, B. & Martínez-Antonio, A., 2011. The Regulatory Network of *Pseudomonas aeruginosa*. *Microbial informatics and experimentation*, 1(1), p.3.

Ganguly, A. & Chatterji, D., 2012. A comparative kinetic and thermodynamic perspective of the σ-competition model in *Escherichia coli*. *Biophysical journal*, 103(6), pp.1325–33.

Geertz, M. & Maerkl, S.J., 2010. Experimental strategies for studying transcription factor-DNA binding specificities. *Briefings in functional genomics*, 9(5-6), pp.362–73.

Gill, S.C., Weitzel, S.E. & von Hippel, P.H., 1991. *Escherichia coli* sigma 70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *Journal of molecular biology*, 220(2), pp.307–24.

Gopalkrishnan, S., Nicoloff, H. & Ades, S.E., 2014. Co-ordinated regulation of the extracytoplasmic stress factor, sigmaE, with other *Escherichia coli* sigma factors by (p)ppGpp and DksA may be achieved by specific regulation of individual holoenzymes. *Molecular Microbiology*, 93(3), pp.479–493.

Govan, J.R. & Fyfe, J.A., 1978. Mucoid *Pseudomonas aeruginosa* and cystic fibrosis: resistance of the mucoid from to carbenicillin, flucloxacillin and tobramycin and the isolation of mucoid variants in vitro. *Journal of Antimicrobial Chemotherapy*, 4(3), pp.233–240.

Grigorova, I.L. et al., 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(14), pp.5332–7.

Gruber, T.M. et al., 2001. Binding of the initiation factor σ70 to core RNA polymerase is a multistep process. *Molecular Cell*, 8(1), pp.21–31.

Gruber, T.M. & Gross, C. a, 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology*, 57, pp.441–466.

Guzman-Vargas, L. & Santillan, M., 2008. Comparative analysis of the complex transcription-factor gene regulatory networks of *E. coli* and *S. cerevisiae*. *BMC systems biology*, 2(1), p.13.

Hamad, B., 2010. The antibiotics market. *Nature reviews. Drug discovery*, 9(9), pp.675–676.

Han, K. et al., 2013. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline milieu. *Scientific reports*, 3, p.2101.

Harrison, F., 2007. Microbial ecology of the cystic fibrosis lung. *Microbiology*, 153(4), pp.917–923.

Haseltine, W.A. & Block, R., 1973. Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 70(5), pp.1564–8.

Helmann, J.D. & Chamberlin, M.J., 1988. Structure and Function of Bacterial Sigma factors. *Ann. Rev. Biochem.*

Hernandez, V.J. & Cashel, M., 1995. Changes in conserved region 3 of *Escherichia coli* sigma 70 mediate ppGpp-dependent functions in vivo. *Journal of molecular biology*, 252(5), pp.536–49.

Hicks, K.A. & Grossman, A.D., 1996. Altering the level and regulation of the major sigma subunit of RNA polymerase affects gene expression and development in *Bacillus subtilis*. *Molecular Microbiology*, 20(1), pp.201–212.

Hindré, T. et al., 2012. New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution. *Nature Reviews Microbiology*, 10(5), pp.352–365.

Horstmann, N. et al., 2011. Distinct single amino acid replacements in the control of virulence regulator protein differentially impact streptococcal pathogenesis. *PLoS Pathogens*, 7(10).

Hughes, K.T. & Mathee, K., 1998. The Anti-Sigma Factors. *Annual Review of Microbiology*, 52(1), pp.231–286.

Jelsbak, L. et al., 2007. Molecular Epidemiology and Dynamics of *Pseudomonas aeruginosa* Populations in Lungs of Cystic Fibrosis Patients. *Infection and Immunity*, 75(0019-9567 (Print) LA - ENG PT - JOURNAL ARTICLE), pp.2214–24.

Jishage, M. et al., 2002. Regulation of sigma factor competition by the alarmone ppGpp. *Genes and Development*, 70, pp.1260–1270.

Karlebach, G. & Shamir, R., 2008. Modelling and analysis of gene regulatory networks. *Nature*

reviews. Molecular cell biology, 9(10), pp.770–780.

Koch, C. & Høiby, N., 1993. Pathogenesis of cystic fibrosis. *The Lancet*, 243, pp.1065–1069.

Kong, W. et al., 2015. ChIP-seq reveals the global regulator AlgR mediating cyclic di-GMP synthesis in *Pseudomonas aeruginosa*. *Nucleic Acids Research*, 43(17), p.gkv747.

Koskella, B. & Vos, M., 2015. Adaptation in Natural Microbial Populations. *Annual Review of Ecology, Evolution, and Systematics*, 46, pp.503–522.

Koskiniemi, S. et al., 2012. Selection-driven gene loss in bacteria. *PLoS Genetics*, 8(6), pp.1–7.

Kulbachinskiy, A. & Mustaev, A., 2006. Region 3.2 of the σ subunit contributes to the binding of the 3´-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *Journal of Biological Chemistry*, 281(27), pp.18273–18276.

Laurie, A.D. et al., 2003. The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. *The Journal of biological chemistry*, 278(3), pp.1494–503.

Leid, J.G. et al., 2005. The exopolysaccharide alginate protects *Pseudomonas aeruginosa* biofilm bacteria from IFN-gamma-mediated macrophage killing. *The Journal of Immunology*, 175(11), pp.7512–7518.

Lesley, S. a & Burgess, R.R., 1989. Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. *Biochemistry*, 28(19), pp.7728–7734.

Lonetto, M., Gribskov, M. & Gross, C., 1992. The Sigma7O Family : Sequence Conservation and Evolutionary Relationships. *Journal of bacteriology*, 174(12), pp.3843–3849.

Lyczak, J.B., Cannon, C.L. & Pier, G.B., 2000. Establishment of *Pseudomonas aeruginosa* infection : lessons from a versatile opportunist. , pp.1051–1060.

Macneil, L.T. & Walhout, A.J.M., 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, 21, pp.645–657.

Maeda, H., Fujita, N. & Ishihama, a, 2000. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic acids research*, 28(18), pp.3497–503.

Magnusson, L.U., Farewell, A. & Nyström, T., 2005. ppGpp: A global regulator in *Escherichia coli*. *Trends in Microbiology*, 13(5), pp.236–242.

Martin, D.W. et al., 1993. Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. *Proc Natl Acad Sci U S A*, 90(18), pp.8377–8381.

Marvig, R.L. et al., 2013. Genome Analysis of a Transmissible Lineage of *Pseudomonas aeruginosa* Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genetics*, 9(9).

Mathee, K., McPherson, C.J. & Ohman, D.E., 1997. Posttranslational control of the algT (algU)-encoded sigma 22 for expression of the alginate regulon in *Pseudomonas aeruginosa* and localization of its antagonist proteins MucA and MucB (AlgN). *Journal of Bacteriology*,

179(11), pp.3711–3720.

Mauri, M. & Klumpp, S., 2014. A Model for Sigma Factor Competition in Bacterial Cells. *PLoS Computational Biology*, 10(10), p.e1003845.

Merhej, V., Georgiades, K. & Raoult, D., 2013. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. *Briefings in Functional Genomics*, 12(4), pp.291–304.

Merrick, M., Gibbins, J. & Toukdarian, A., 1987. The Nucleotide-Sequence of the Sigma Factor Gene Ntra (rpon) of *Azotobacter-Vinelandii* - Analysis of Conserved Sequences in Ntra Proteins. *Molecular & General Genetics*, 210(2), pp.323–330.

Mooney, R.A., Darst, S. a & Landick, R., 2005. Sigma and RNA polymerase: an on-again, off-again relationship? *Molecular cell*, 20(3), pp.335–45.

Mowat, E. et al., 2011. *Pseudomonas aeruginosa* population diversity and turnover in cystic fibrosis chronic infections. *American Journal of Respiratory and Critical Care Medicine*, 183(12), pp.1674–1679.

Munck, C. et al., 2014. Prediction of resistance development against drug combinations by collateral responses to component drugs. *Science Translational Medicine*, 6(262), p.262ra156.

Myers, K.S. et al., 2015. Defining bacterial regulons using ChIP-seq. *Methods*, 86, pp.80–88.

Nyström, T., 2004. Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Molecular microbiology*, 54(4), pp.855–62.

Österberg, S., Peso-Santos, T. Del & Shingler, V., 2011. Regulation of Alternative Sigma Factor Use. *Annual Review of Microbiology*, 65(1), pp.37–55.

Paget, M., 2015. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*, 5(3), pp.1245–1265.

Paul, B.J., Berkmen, M.B. & Gourse, R.L., 2005. DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences*, 102(22), pp.7823–7828.

Philippe, N. et al., 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *BioEssays*, 29(9), pp.846–860.

Pieren, M. & Tigges, M., 2012. Adjuvant strategies for potentiation of antibiotics to overcome antimicrobial resistance. *Current Opinion in Pharmacology*, 12(5), pp.551–555.

Poole, K., 2001. Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. *Journal of Microbiol. Biotechnol.*, 3(2), pp.255–264.

Potrykus, K. & Cashel, M., 2008. (p)ppGpp: Still Magical? *. *Annual Review of Microbiology*, 62(1), pp.35–51.

Price, M.N., Dehal, P.S. & Arkin, A.P., 2007. Orthologous transcription factors in bacteria have different functions and regulate different genes. *PLoS Computational Biology*, 3(9), pp.1739–1750.

Pupov, D. et al., 2014. Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic acids research*, 42(7), pp.4494–504.

Qiu, D. et al., 2007. Regulated proteolysis controls mucoid conversion in P*seudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), pp.8107–12.

Raffaelle, M. et al., 2005. Holoenzyme switching and stochastic release of sigma factors from RNA polymerase *in vivo*. *Molecular cell*, 20(3), pp.357–66.

Rau, M.H. et al., 2012. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environmental Microbiology*, 14(8), pp.2200–2211.

Renzoni, A. et al., 2011. Whole genome sequencing and complete genetic analysis reveals novel pathways to glycopeptide resistance in staphylococcus aureus. *PLoS ONE*, 6(6).

Rollenhagen, C. et al., 2003. Binding of σ A and σ B to Core RNA Polymerase after Environmental Stress in *Bacillus subtilis Journal of bacteriology*, 185, pp.35–40.

Ross, W. et al., 2013. The magic spot: A ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Molecular Cell*, 50(3), pp.420–429.

Rothschild, L.J. & Mancinelli, R.L., 2001. Life in Extreme Environments. *Nature*, 409(February 2001), pp.1092–1101.

Schulz, S. et al., 2015. Elucidation of Sigma Factor-Associated Networks in *Pseudomonas aeruginosa* Reveals a Modular Architecture with Limited and Function-Specific Crosstalk. *PLOS Pathogens*, 11(3), p.e1004744.

Schurr, M.J. et al., 1996. Control of AlgU, a member of the σ(E)-like family of stress sigma factors, by the negative regulators mucA and mucB and *Pseudomonas aeruginosa* conversion to mucoidy in cystic fibrosis. *Journal of Bacteriology*, 178(16), pp.4997–5004.

Schurr, M.J. et al., 1993. The algD promoter: regulation of alginate production by *Pseudomonas aeruginosa* in cystic fibrosis. *Cell Mol Biol Res*, pp.371–376.

Severinov, K. et al., 1994. The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *The Journal of biological chemistry.*, 269(33), pp.20826–20828.

Sharp, M.M. et al., 1999. The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes and Development*, 13(22), pp.3015–3026.

Shawar, R.M. et al., 1999. Activities of tobramycin and six other antibiotics against *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. *Antimicrobial agents and chemotherapy*, 43(12), pp.2877–2880.

Shingler, V., 2011. Signal sensory systems that impact sigma 54 -dependent transcription. *FEMS Microbiology Reviews*, 35, pp.425–440.

Smith, E.E. et al., 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic

fibrosis patients. *Proc.Natl.Acad.Sci.U.S.A*, 103(22), pp.8487–8492.

Sousa, A.M. & Pereira, M.O., 2014. *Pseudomonas aeruginosa* Diversification during Infection Development in Cystic Fibrosis Lungs-A Review. *Pathogens (Basel, Switzerland)*, 3(3), pp.680–703.

Studholme, D.J. & Dixon, R., 2003. Domain Architectures of σ 54 -Dependent Transcriptional Activators. *Journal of bacteriology*, 185(6), pp.1757–1769.

Taylor, M. et al., 1996. The RpoN-box motif of the RNA polymerase sigma factor sigma N plays a role in promoter recognition. *Molecular microbiology*, 22(5), pp.1045–54.

Thomassen, M.J. et al., 1979. Multiple of isolates of *Pseudomonas aeruginosa* with differing antimicrobial susceptibility patterns from patients with cystic fibrosis. *The Journal of infectious diseases*, 140(6), pp.873–880.

Tomich, M., Planet, P.J. & Figurski, D.H., 2007. The *tad* locus : postcards from the widespread colonization island. *Nature Reviews Microbiology*, 5(May), pp.363–375.

Welsh, M.J. & Smith, a E., 1993. Molecular mechanisms of CFTR Chloride Channel Dysfunction in Cystic Fibrosis. *Cell*, 73(7), pp.1251–1254.

Wiedenbeck, J. & Cohan, F.M., 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, 35(5), pp.957–976.

World Health Organisation, 2015. Antimicrobial Resistance. *Fact Sheet No 194*.

Xie, Z.D. et al., 1996. Sigma factor-anti-sigma factor interaction in alginate synthesis: inhibition of AlgT by MucA. *Journal of bacteriology*, 178(16), pp.4990–4996.

Yang, L., Jelsbak, L., Marvig, R.L., et al., 2011. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), pp.7481–6.

Yang, L., Jelsbak, L. & Molin, S., 2011. Microbial ecology and adaptation in cystic fibrosis airways. *Environmental Microbiology*, 13(7), pp.1682–1689.

Yang, Y. et al., 2015. Structures of the RNA polymerase-sigma54 reveal new and conserved regulatory strategies. *Science*, 349(6250), pp.882–886.

Yin, Y. et al., 2013. Evidence for sigma factor competition in the regulation of alginate production by *Pseudomonas aeruginosa*. *PloS one*, 8(8), p.e72329.

Zhou, Y.N. & Gross, C., 1992. How a mutation in the gene encoding sigma 70 suppresses the defective heat shock response caused by a mutation in the gene encoding sigma 32. *Journal of bacteriology*, 174(22), pp.7128–7137.

Zhou, Y.N., Walter, W. a & Gross, C., 1992. A mutant sigma 32 with a small deletion in conserved region 3 of sigma has reduced affinity for core RNA polymerase. *Journal of bacteriology*, 174(15), pp.5005–12.

Zuo, Y., Wang, Y. & Steitz, T.A., 2013. The Mechanism of *E. coli* RNA Polymerase Regulation by

ppGpp is suggested by the structure of their complex. *Molecular Cell*, 50(3), pp.430–436.

Chapter 7

Research articles

The research articles are enclosed in the following order:

Paper 1

Eva Kammer Andresen, Denitsa Eckweiler, Sebastian Schulz, Susanne Haussler, Tino Krell Maher Abou Hachem, Lars Jelsbak (2016). **Rewiring of a sigma factor regulatory network in *P. aeruginosa* by a naturally occurring single nucleotide polymorphism**. Manuscript submitted for publication.

Paper 2

Eva Kammer Andresen, Denitsa Eckweiler, Sebastian Schulz, Grith Mirriam Maigaard Hermansen, Susanne Haussler, Maher Abou Hachem, Lars Jelsbak (2016). **Epistasis and sigma factor competition in *Pseudomonas aeruginosa.*** Manuscript in preperation.

Rewiring of a sigma factor regulatory network in *P. aeruginosa* by a naturally occurring single nucleotide polymorphism

Eva Kammer Andresen[1], Denitsa Eckweiler[2], Sebastian Schulz[2], Susanne Haussler[2], Tino Krell[3], Maher Abou Hachem[1], Lars Jelsbak[1*]

[1] Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark [2] Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany. [3]Department of Environmental Protection, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, C/ Prof. Albareda, 1, 18008, Granada, Spain

*Corresponding author: DTU Systems Biology, Building 301, 2800 Kgs. Lyngby, Denmark, lj@bio.dtu.dk, +4545256129

Keywords: *Pseudomonas aeruginosa*, evolution, transcriptional regulation, sigma factor, σ^{54}, promoter recognition

Abstract

Transcriptional regulatory networks represent an important organisational element in the bacterial cell where signals from the cell state and the outside environment are integrated in terms of activation and inhibition of gene transcription. How these networks evolve is not well understood, and the effects of natural occurring polymorphisms within network components are often not known. Here, we systematically investigate the functional consequences of an amino acid substitution in the sigma factor protein RpoN in the opportunistic pathogen *Pseudomonas aeruginosa*. While RpoN is known to be important for virulence gene expression in *P. aeruginosa*, the particular amino acid substitution is important for ecological success of the bacteria in cystic fibrosis infections. We use Chromatin Immunoprecipitation coupled with next generation sequencing (ChIP-seq), transcriptional profiling, as well as *in vitro* protein-DNA interaction studies and show that the *rpoN* mutation results in reduced connections to only parts of the RpoN controlled regulatory network. On the other hand, the mutation results in activation of other sigma factor regulons. Finally, our results also show that the *rpoN* mutation lead to a rewiring of the RpoN network in terms of increased connectivity and positive regulatory effect on a virulence associated locus. This molecular pleiotropy could not have been predicted from *in silico* analyses alone, and this work thus underlines the importance of achieving a molecular understanding of the effects of polymorphisms in regulator network components.

Introduction

Coordinated global transcriptional changes during infection are critical to the pathogenesis of most bacteria that infect humans, and transcriptional regulatory networks of connected transcriptional regulators and target genes play a central role in this process by transducing signals from the physiological state of the cell or the surrounding environment into coordinated expressions of the genome. At the molecular level, the structure and function of many transcriptional regulatory networks are well understood (Salgado et al. 2013; Rustad et al. 2014; Schulz et al. 2015). In contrast, it is less understood how transcriptional regulatory networks evolve and how ecological and evolutionary forces act to maintain or diversify these networks, and in which way evolutionary changes in regulatory networks influence bacterial pathogenesis.

Recent studies have pointed out that evolution of pathogenic potential and host adaptation may involve rewiring of pre-existing regulatory networks in the pathogen genome. For example, point mutations within promoter regions can create novel regulatory connections critical for pathogen fitness within hosts (Osborne et al. 2009) and non-synonymous mutations in global regulator genes can lead to loss of regulatory network connections resulting in reduced virulence of the pathogen (D'Argenio et al. 2007; Horstmann et al. 2011; Flores et al. 2015). In addition to these specific cases, advances in genome sequencing of bacterial pathogens have documented regulatory protein sequence diversity among strains of the same species (Shea et al. 2011). In some cases, this genetic diversity is selected by evolution during the course of infection in individual patients (Lieberman et al. 2011; Marvig et al. 2013). However, interpretation of the molecular effects of this genetic diversity is often difficult since knowledge about gene functionality is most often derived from experimentation with genetically engineered loss-of-function mutants and not the exact polymorphisms selected during the course of evolution. So, while these observations clearly indicate that genetic changes in regulatory networks play an important role in connection with

pathogen adaptation and evolution, the precise functional consequences of genetic variations within regulatory components are most often not known.

Among the different families of transcriptional regulator proteins, sigma factors are of exceptional importance in connection with coordination of global gene expression as they provide promoter recognition specificity by directing the catalytic core RNA Polymerase to different promoter sequences (Burgess & Travers 1969). The number of sigma factors varies greatly between bacterial species, from only 1 in *Mycoplasma* sp. to more than 100 in the *Sorangium cellulosum* (Gruber & Gross 2003; Han et al. 2013). The opportunistic pathogen *Pseudomonas aeruginosa* encodes more than 25 sigma factors, most of which have been linked to bacterial virulence and pathogenicity (Potvin et al. 2008). In *P. aeruginosa*, one of these sigma factors (RpoN) has been shown to regulate multiple cellular functions important for pathogenicity, including alginate production (Boucher et al. 2000), assembly of motility organs (Dasgupta et al. 2003), utilisation of alternative carbon sources (Valentini et al. 2011), and production of virulence factors such as secretion systems (Dong & Mekalanos 2012). In spite of the many regulatory effects associated with RpoN, *P. aeruginosa* strains from chronic airway infections in individuals with the genetic disorder cystic fibrosis (CF) have repeatedly been observed to acquire non-synonymous mutations in *rpoN* during host colonisation (Smith et al. 2006). We have recently documented the fixation of a specific *rpoN* mutation ($rpoN^{DK2}$) in the epidemic *P. aeruginosa* DK2 clone type during its decade long adaptation to the CF airway environment and have shown that the mutation contributed to the ecological success of the DK2 lineage and the transformation of the lineage from an opportunistic pathogen to a CF-specific pathogen (Damkiaer et al. 2013). Although *rpoN* mutations are selected by evolution in CF airways, the specific molecular effects of these mutations on the function and activity of RpoN remain unknown.

Here, we systematically investigate the functional consequences of the DK2 specific amino acid substitution in RpoN using a combination of Chromatin Immunoprecipitation coupled with next generation sequencing (ChIP-seq), transcriptional profiling, as well as *in vitro* protein-DNA interaction studies. We show that the *rpoN*^{DK2} mutation results in reduced connections to only parts of the RpoN controlled regulatory network and enhanced crosstalk effects to other sigma factor regulons. Importantly, our results show that the *rpoN*^{DK} mutation resulted in an increased connectivity and positive regulatory effect on the *tad* locus, which is involved in biofilm formation, colonisation, and pathogenesis in a wide range of bacterial species (Tomich et al. 2007). This work not only provides a molecular explanation of a naturally occurring global regulator mutation, it also highlights the need for in-depth molecular analyses of specific mutational events, as these may result in unpredictable transcriptional rewiring and phenotypic changes.

Materials and methods

Genetic manipulations

DNA extraction, restriction enzyme treatments, ligation of DNA fragments, and transformation of *Eschericia coli* were performed using standard methods (Green & Sambrook 2012).

Chromatin immunoprecipitation

ChIP-seq analysis of RpoN^{WT} was performed by transforming plasmid pJN105-*rpoN*^{WT} (Schulz et al. 2015) into PAO1. Construction of pJN105-*rpoN*^{DK2} was performed with the Quick Change Lightning Site Directed Mutagenesis kit (Stratagene) with the primers SD6-F and SD6-R (Table S1). The resulting plasmid pJN105-*rpoN*^{DK2} was then transformed into PAO1$^{rpoN(DK2)}$ (Damkiaer et al. 2013).

ChIP experiments were performed as described previously (Schulz et al. 2015). Briefly, 50ml cultures were grown to an OD_{600} of 1.0, and sigma factor expression was induced with l-arabinose to 0.5% (w/v). Crosslinking was performed by adding 0.5% formaldehyde when OD_{600} reached 2.0. Before formaldehyde treatment, 1.5 ml culture was sampled for mRNA profiling. The crosslinking reaction was quenched after 5 minutes with 137 mM glycine, and the cells were pelleted. Pellets were resuspended in lysis buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 10 mM EDTA, 20% sucrose), and DNA was fragmented to an average size of 200-250 bp, using the method of (Schulz et al. 2015). Bound DNA was incubated with anti-6xHistag antibody (ab9108, Abcam), and antibody-protein-DNA complexes were captured using Dynabeads Protein G (Life Technologies). Immunoprecipitated samples were treated with RNAse A and proteinase K, and DNA was purified using the QIAquick PCR purification Kit (Qiagen). Recovered DNA was subjected to a modified linear DNA amplification (LinDA) protocol (Shankaranarayanan et al. 2012; Schulz et al. 2015), and preparation of libraries for sequencing was performed with the NEBNext Ultra DNA Library Prep Kit for Illumina. DNA was sequenced as paired-end sequences with a read length of 250 bp on a MiSeq sequencer (Illumina).

ChIP-seq data was analysed in a way similar to that described previously (Schulz et al. 2015). Because of the paired-end sequencing and read length, we chose the Stampy aligner (Lunter & Goodson 2011) to map the reads to the PAO1 reference genome. For both $RpoN^{WT}$ and $RpoN^{DK2}$, model-based analysis of ChIP-seq (MACS) (Zhang et al. 2008), was applied for peak detection using a P value cut-off value of 0.05 and shift size 30 for the peak modelling. Promoter hits [-500nt, TSS, +100nt] were considered significant when they were detected in both ChIP-seq replicates with an enrichment factor of at least 2 and a P value of less than 0.01. Statistical analysis of the obtained candidates was performed to assess the number of false positives and the corresponding P value according to the hypergeometric test in R using the phyper command.

Gene expression analysis

The 1.5 ml samples collected above were used for isolation of total RNA using the RNeasy Mini Kit (Qiagen), and enrichment of mRNA was performed using the ScriptSeq Complete Kit (Bacteria). cDNA library preparation was performed as described previously (Schulz et al. 2015), and cDNA was sequenced as paired-end sequences with a read length of 250 bp on a MiSeq sequencer (Illumina).

Sequences were mapped to the PAO1 genome using Stampy (Lunter & Goodson 2011). Differential expression was called with the R package DESeq2 (Love et al. 2014) making use of the biological duplicates for each condition. As differentially expressed genes were identified, those having absolute logarithmic fold change greater than one ($|\log_2 FC|>1$) at significance level of less than 5%.

Expression and purification of recombinant proteins

The $rpoN^{WT}$ and $rpoN^{DK2}$ genes were amplified from genomic DNA prepared from *P. aeruginosa* PAO1 and PAO1$^{rpoN(DK2)}$ (Damkiaer et al. 2013), respectively. Both genes were amplified with the primers rpoN_F and rpoN_R (Table S1) and ligated into plasmid pET28 (Novagen, Merck Millipore) using restriction sites NdeI and HindIII, producing N-terminal 6xHistidine tagged sequences of each insert. All plasmids were verified by Sanger sequencing and maintained in *E. coli* DH5α.

The recombinant plasmids generated above were transformed into the expression *E. coli* Rosetta(DE3) strain (Novagen, Merck Millipore). Cells were grown in LB medium supplemented with 10 mM glucose, 50 µg/ml kanamycin, and 34 µg/ml chloramphenicol in 1L cultures at 30°C to an OD_{600} of 0.5. Expression was induced by adding IPTG to a final concentration of 100 µM, and the cultures were harvested after 3 hours and stored at -20°C until use.

Purification was performed by resuspending 5 g frozen cells in 30 ml buffer A (10 mM HEPES pH 7.5, 15 mM Imidazole, 0.5M NaCl, and 10% Glycerol), supplemented with 7µl Benzonase (Sigma-Aldrich) and 1 cOmplete protease inhibitor tablet (Roche Diagnostics) and passed through a French Press 3 times at 1,000 bar. Cell lysates were centrifuged, and the clarified supernatant was loaded onto a 5 ml HisTrap column (GE Healthcare) at 0.5 ml/min and 4˚C, washed with 6 column volumes (CV) of the same, after which bound fractions were eluted with buffer B (10 mM HEPES pH 7.5, 400 mM Imidazole, 0.5M NaCl, and 10% Glycerol) in 30 CV. Fractions containing 90%< pure $RpoN^{WT}$ and $RpoN^{DK2}$ were collected and dialysed against storage buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), and 10% glycerol). Proteins were stored at -80°C in storage buffer and glycerol to a final concentration of 50%.

Preparations of DNA for RpoN – DNA interaction analysis

Duplex DNA oligos containing RpoN promoter recognition sequences for *fliD* and *flgF* (enclosed in Table S1) were purchased from Integrated DNA Technologies. Promoter duplexes were designed from the core sequence of the known RpoN binding motif GG-N_{10}-GC (Taylor et al. 1996; Buck et al. 2000), and in addition to this, an additional 15 bp spacer was added to each end creating a final duplex consisting of N_{15}-GG-N_{10}-GC-N_{15},where N denotes the genomic DNA between the -24 GG and -12 GC nucleotides as well as 15 bp up- and downstream of these. The 305 bp intergenic region between genes *rcpC* and *flp* was amplified from *P. aeruginosa* PAO1 genomic DNA using primers Intreg_F and Intreg_R.

Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSA) were performed using RpoN promoter sequences end-labelled with γ-33-ATP using T4 polynucleotide kinase (Thermo Fischer Scientific) and purified using Micro Bio-Spin 6 Chromatography Columns (Bio-Rad Laboratories). EMSA

reactions were assembled in 10 µl reactions in binding buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM DTT, and 10% glycerol) using 1 µl of labelled DNA diluted to $1.5 \cdot 10^4$ cpm, 0.2 µl 50 mg/ml Bovine Serum Albumin (BSA), 0.5 µl 1 mg/ml Poly[d(I-C)], and increasing concentrations of purified $RpoN^{WT}$ or $RpoN^{DK2}$. Reactions were incubated for 20 minutes at 30°C, after which 1 µl Novex® Hi-Density TBE Sample Buffer (5X) (ThermoFischer Scientific) was added, and the electrophoresis was performed using Novex® TBE Gels 6% DNA Retardation Gels (ThermoFischer Scientific) and visualised on a Storm PhosphorImager (GE Healthcare).

Results

Sequence variation of the *Pseudomonas aeruginosa* σ^{54} RpoN

To investigate the natural sequence variation of RpoN, we performed a BLAST search of the PAO1 RpoN amino acid sequence against the entire set of *P. aeruginosa* RpoN sequences deposited in the *Pseudomonas* Genome Database (Winsor et al. 2016). We found that 10.8% (163 of 1,507 complete RpoN sequences) display sequence variation of one or more amino acid substitutions (Dataset S1). One of these is the $RpoN^{DK2}$ protein from the transmissible *P. aeruginosa* clone DK2, originally isolated from chronically infected Danish CF patients. The $RpoN^{DK2}$ protein has a single amino acid substitution, L419P, which is known to result in phenotypic changes and affect regulatory networks (Damkiaer et al. 2013).

The L419P substitution maps to the DNA interacting domain of RpoN

To infer about the possible consequences of the $RpoN^{DK2}$ mutation, we performed a sequence alignment between *E. coli* RpoN and *P. aeruginosa* RpoN. The $RpoN^{DK2}$ L419P substitution maps to region 3, which is known to be involved in DNA interactions (Buck et al. 2000; Österberg et al. 2011) (**Figure 1**). Region 3 of RpoN contains 3 domains; a core binding domain (CBD), a helix-turn-helix (HTH) motif involved in interaction with the -12 promoter region, and the RpoN-box

involved in interactions with the -24 promoter sequence (Buck et al. 2000; Österberg et al. 2011). The RpoN^{DK2} mutation is located between the HTH and RpoN-box, and we therefore hypothesised that the mutation could potentially affect promoter DNA binding abilities, and thus, the ability of RpoN^{DK2} to control its regulon.

Regulatory effects of the RpoN^{DK2} mutation

To determine the regulatory effects of the RpoN^{DK2} mutations, we first examined the transcriptional profile of PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} compared to PAO1$^{rpoN(WT)}$ expressing RpoN^{WT} to create an overall assessment of both direct and indirect effects of the RpoN^{DK2} mutation (**Figure 2**, grey series) (Dataset S2). At a significance level of $p<0.05$ and log fold change ($|\log_2 FC|>1$, 454 genes were significantly differentially expressed, with 341 upregulated and 113 downregulated genes, altogether corresponding to 8.1% of the total 5,570 open reading frames in *P. aeruginosa* PAO1 (Stover et al. 2000).

We next compared our gene expression values to the known core RpoN regulon (defined as genes containing promoters known to interact directly with RpoN (Schulz et al. 2015) to investigate if the RpoN^{DK2} mutation results in a specific up- or downregulation of genes known to be controlled by RpoN. This analysis shows that genes directly controlled by RpoN are mostly downregulated (47 genes of a total of 66 belonging to the core RpoN regulon, corresponding to 71% of the total core RpoN regulon) (**Figure 2**, red series). These results indicate that the RpoN^{DK2} mutation has a negative impact on the RpoN regulon as well as substantial pleotropic effects outside the RpoN regulon.

The RpoN^{DK2} mutation positively affects the transcription profile of the RpoS regulon

To further understand these pleotropic effects, we explored whether any of the other sigma factor core regulons characterised in *P. aeruginosa* were affected by the RpoN^{DK2} mutation. To this end,

we compared the total expression dataset from the PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} compared to PAO1$^{rpoN(WT)}$ expressing RpoN^{WT} to six of the known sigma factor core regulons (RpoD, PvdS, AlgT, RpoS, FliA, RpoH) (**Figure 3**). We did not observe any changes in relation to the genes from the core regulons of RpoD, PvdS, FliA, and RpoH (Schulz et al. 2015). The differential expression values of the genes belonging to the core regulons of AlgT seemed to be slightly downregulated. On the other hand, the entire set of genes that constitute the RpoS regulon was upregulated, suggesting the existence of interplay between the RpoN and RpoS regulons. Data for all comparisons are provided as Dataset S3.

Genome wide binding site distributions of the RpoN^{WT} and RpoN^{DK2} proteins

To separate the direct and indirect regulatory effects of the RpoN^{DK2} mutation, and to investigate if the RpoN^{DK2} mutation affects the genome wide binding profile of the sigma factor protein, we performed a ChIP-seq analysis, and compared the binding profiles of PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} to PAO1$^{rpoN(WT)}$ expressing RpoN^{WT} (Dataset S4 and S5). This analysis showed that a majority of unique binding sites were associated with RpoN^{WT} (286 binding sites), whereas three binding sites appeared to be responsive to RpoN^{DK2}, but not the RpoN^{WT} protein (**Figure 4A**).

The gene expression analysis and genome wide binding profiling of the RpoN^{DK2} mutation show that the direct effects of the RpoN^{DK2} mutation is a reduced number of binding sites of the RpoN^{DK2} protein compared to the RpoN^{WT} protein, as well as reduced expression levels of genes that constitute the core RpoN regulon, indicating a reduced binding affinity of the RpoN^{DK2} protein to its promoters. The gene expression profile, however, also reveals an RpoN^{DK2} positive regulation of a number of genes, indicating a more complex regulatory rewiring than merely a loss-of-function.

The RpoN^{DK2} mutation decreases the direct binding affinity to RpoN promoter sequences

To verify if the RpoN^{DK2} protein displays a reduced binding affinity to RpoN controlled promoters, we used EMSA to compare the *in vitro* binding profile of RpoN^{WT} and RpoN^{DK2} recombinant proteins to two well-known RpoN controlled promoters, both called as unique RpoN^{WT} binding sites in our ChIP-seq experiments. Figure 4B shows the binding of each protein to these two promoters. It is clear that at comparable concentrations of the RpoN^{WT} and RpoN^{DK2} protein (marked by asterisks), the RpoN^{WT} protein produces a stronger binding at both promoters, meaning that the binding affinity of the RpoN^{DK2} protein is decreased compared to the RpoN^{WT}. It was not possible to determine absolute binding affinities due to instability of the purified RpoN proteins at high concentration in solution, however, it is clear that the binding of RpoN^{DK2} is either reduced or lost, confirming the *in vivo* observations from ChIP-seq.

Remodelling of the RpoN^{DK2} regulon

To investigate the unique RpoN^{DK2} binding events discovered from the ChIP-seq analysis, we correlated the three unique RpoN^{DK2} ChIP peaks with their corresponding mRNA values and discovered a clear association between the RpoN^{DK2} pull-down of two gene promoter regions (PA4305, *rcpC* and PA4306, *flp*), and upregulation of their gene expression values (**Figure 6A**). The *flp* and *rcpC* genes are transcribed in opposite directions, and both are members of the *tad* locus, responsible for the production and assembly of type IVb pili (Bernard et al. 2009). Interestingly, the positive regulatory effects of RpoN^{DK2} are not limited to *flp* and *rcpC*. As shown in Figure 6B, most genes of the *tad* locus are upregulated in response to the RpoN^{DK2} mutation, although *flp* and *rcpC* are the only two genes producing unique ChIP binding sites for the RpoN^{DK2} protein.

We next visually inspected the ChIP binding profiles of $RpoN^{WT}$ and $RpoN^{DK2}$ at the 1,285 bp region surrounding the genes *rcpC*, *flp*, and *pctC*, in order to evaluate the actual binding profile of each protein to this region (**Figure 7A**). Figure 7.A.1 - 7.A.2 illustrate the replicates of the ChIP binding profile of the $RpoN^{WT}$ protein, whereas Figure 7.A.3 - 7.A.4 show the replicates of the ChIP binding profile of the $RpoN^{DK2}$ protein at this region. While bioinformatics analyses call the peak in this region as being unique for the $RpoN^{DK2}$, visual inspection indicates that minimal binding of the $RpoN^{WT}$ to this region is possible.

To confirm any role of direct $RpoN^{DK2}$ binding to the intergenic region of *flp-rcpC*, we searched this intergenic sequence for RpoN consensus sites. A putative RpoN consensus site was discovered at the genomic coordinates 4830896-4830912 with the sequence 5´-TCGGCCTAGCCTCAGCG-3´. Figure 7B shows *in vitro* EMSA binding ability of both $RpoN^{WT}$ and $RpoN^{DK2}$ to a 305 bp region of the intergenic region between *rcpC* and *flp* containing the putative RpoN consensus sequence. Even though both $RpoN^{WT}$ and $RpoN^{DK2}$ are able to bind this region *in vitro*, the $RpoN^{DK2}$ protein produces a complete shift at 1.1 µM, while $RpoN^{WT}$ requires 4.7 µM - 6.3 µM to produce a complete shift, indicating that the $RpoN^{DK2}$ protein binds this region with a higher affinity.

Discussion

The increased use of high throughput comparative genomics and *in silico* modelling of biological systems presents both fascinating possibilities as well as complex challenges when used to model complicated networks such as transcriptional regulatory networks. However, in order to achieve an in-depth genotype-phenotype understanding and an ability to predict resulting phenotypes from genomic changes, there is a need for understanding the molecular consequences of genotypic alterations e.g. mutations in regulatory proteins.

In this work, we set out to investigate how one naturally occurring amino acid substitution of the *P. aeruginosa* sigma factor RpoN alters the transcriptional profile, genome wide binding profile, and the *in vitro* DNA binding to the RpoN recognised promoter DNA. We performed experiments both *in vivo* and *in vitro,* as this combination represents a powerful combination of techniques. Whereas *in vitro* EMSA binding presents the direct effect of a mutation on a defined DNA sequence, results of *in vivo* experiments reveal the entire biological response of the mutation.

We initially hypothesised that the $RpoN^{DK2}$ mutation would result in an overall reduced binding affinity to RpoN controlled promoters, reflected in a consistent downregulation of genes belonging to the RpoN regulon, a reduced genome wide RpoN binding site distribution, as well as a decrease in binding affinity of $RpoN^{DK2}$ promoters on *in vitro* EMSA binding profiles. Interestingly, we discovered that the mutation results in multiple complex regulatory effects, including a decreased binding of the $RpoN^{DK2}$ protein to certain RpoN controlled promoters, as well as a remodelling of the entire RpoN regulon. In addition, we discovered a unique positive regulatory effect of the $RpoN^{DK2}$ protein on the *tad* locus, a region that has not previously been linked to direct RpoN regulation.

The random distribution of expression values of genes known to be members of the RpoN regulon suggests that the $RpoN^{DK2}$ mutation is not comparable to a loss-of-function mutant, but rather results in an altered function. A similar situation has been observed in the group A *Streptococcus,* where a naturally occurring mutation in the sensor kinase LiaS resulted in an alteration, but not elimination of the LiaS protein function (Flores et al. 2015).

Interestingly, among the upregulated genes in our gene expression analysis was the entire set of genes constituting the core RpoS regulon, indicating an additional regulatory effect on sigma factor crosstalk. Sigma factor crosstalk has been proposed to be limited, but highly function-specific, and

315 RpoN has been proposed as the sigma factor participating in the most extensive crosstalk among
316 sigma factors with direct crosstalk to AlgT, FliA, SigX, RpoH, and RpoS (Schulz et al. 2015). Our
317 findings that the $RpoN^{DK2}$ mutation results in an RpoS specific crosstalk, but leaves other regulons
318 unaffected, support the hypothesis that the $RpoN^{DK2}$ mutation is a regulon modulating mutation that
319 is able to direct cellular processes towards specific functions.

320 Specific interactions between the RpoN and RpoS regulatory networks have been reported
321 previously. In *Borrelia burgdorferi*, a central regulatory pathway, known as the σ^{54}-σ^{S} sigma factor
322 cascade, controls expression of the surface protein OspC and a large number of other genes, and it
323 has been shown for this organism that RpoN directly controls RpoS expression (Smith et al. 2007;
324 Ouyang et al. 2012). However, it has also been suggested that this specific σ^{54}-σ^{S} sigma factor
325 cascade requires an activator protein that activates transcription through a DNA binding mechanism
326 different from the normal RpoN activation mechanism, and is dependent on direct repeat sequences
327 upstream of the *rpoS* gene (Ouyang et al. 2011). Our findings that $RpoN^{DK2}$ positively affects RpoS
328 expression may be yet another variation of the RpoN-RpoS regulatory interplay that is dependent on
329 the genetic environment and activator proteins.

330 The $RpoN^{DK2}$ mutation was originally observed in the *P. aeruginosa* DK2 lineage that has
331 successfully adapted to CF hosts over more than 40 years (Yang et al. 2011). In addition to the
332 $RpoN^{DK2}$ allele discussed here, many other mutations in regulatory proteins have been observed in
333 the DK2 lineage (Damkiaer et al. 2013), and it appears that regulatory mutations in addition to
334 acquisition and loss of genetic material have all been driving factors in the evolution of this specific
335 lineage. While this study determines the direct regulatory and molecular effects of the $RpoN^{DK2}$
336 mutation in a PAO1 background, a complete evolutionary understanding of this mutation, as well as
337 all other adaptive regulatory mutations should be discussed in relation to their genomic
338 backgrounds. It is very likely that compensatory mutations, as well as new or lost genetic material,

will add another layer of complexity, especially when taking into consideration that the RpoN sigma factor interacts not only with promoter DNA, but also with the core RNAP, bEBPs, and that transcriptional activation of RpoN requires correct bEBP binding to upstream DNA sequences, as well as correct DNA looping facilitated by an integration host factor (IHF) protein. In fact, it was recently reported that mutations in two *Pseudomonas flourescens* bEBPs result in pleiotropic effects and unpredictable phenotypes, and activate transcription from certain promoters (Taylor, Mulley, Mcguffin, et al. 2015; Taylor, Mulley, Dills, et al. 2015). Using the same set of bEBPs (NtrC/NtrB), Amit et al. 2011 showed that transcriptional control from these two regulators depends quantitatively on DNA looping. Taken together, these studies of mutations in bEBP, the importance of the surrounding genetic environment, as well as our findings of an $RpoN^{DK2}$ facilitated remodelling of regulatory networks, all point towards an extremely complex system that is accessible for regulation on many levels.

The *tad* locus presents a specific example of the complexity of regulatory networks. A literature search did not reveal any previously described direct RpoN regulation of *flp* and *rcpC* genes, and we therefore speculate that this locus may present a new regulatory function of the $RpoN^{DK2}$ protein. The improved binding ability of $RpoN^{DK2}$ at this region is directly translated into increased gene expression for the *rcpC* and *flp* genes, indicating that the specific interaction takes place in the intergenic region between *rcpC* and *flp*, and that any altered regulatory effects exercised by the $RpoN^{DK2}$ are specific for *rcpC* and *flp*, but not *pctC*. This interesting case shows that while the $RpoN^{DK2}$ protein seems to lose some binding sites, based on the high number of unique binding events for the $RpoN^{WT}$ protein, the mutation also results in an altered regulatory function at certain genomic regions.

The potential $RpoN^{DK2}$ facilitated regulatory mechanism of this specific locus may be a trait that has been preserved in the evolved *P. aeruginosa* DK2 strain, or it could represent a trait with no

evolutionary beneficial value that was later silenced by compensatory mutations. Indeed, sequence alignment of the genes constituting the *tad* locus in *P. aeruginosa* PAO1 and DK2 reveals a number of mutations in the *tad* locus genes (**Table S2**). So far, the only in-depth descriptions of the *P. aeruginosa tad* locus have been those of (Bentzmann et al. 2006) and (Bernard et al. 2009). The genes comprising the *tad* locus have furthermore been proposed to be directly regulated by RpoS, FliA, and RpoH. Indirect regulation of the *tad* locus involves RpoN, FliA, RpoS, and RpoH (Schulz et al. 2015). The aforementioned roles of the surrounding genetic environment, and both bEBPs and RpoN as additional evolutionary targets, represent an interesting starting point for future investigations.

In summary, our results reporting that a single amino acid substitution in a well-known sigma factor leads to extensive remodelling of its regulatory network as well as an altered DNA binding ability are a striking example of the difficulties one is confronted with when studying the evolution of regulatory networks. Our findings not only elucidate the specific molecular mechanisms of this mutation, they also highlight the importance of understanding evolution of regulatory networks in their specific context. Ultimately, this should lead to a closer examination of how evolution of regulatory networks varies within specific strains, and to which extent information on one strain is applicable to even closely related strains.

Acknowledgements

We would like to thank Susanne Kofoed and Tanja Nicolai for expert technical assistance, and Agata Bielecka for assistance on ChIP- and mRNA experiments. We would also like to thank the Augustinus Foundation and Oticon Foundation for travel grants, as well as the Villum Foundation for funding for this study to LJ (Grant number VKR023113). LJ acknowledges additional funding from the Novo Nordisk Foundation and the Lundbeck Foundation.

Competing interests

None

Reference list

Amit, R. et al., 2011. Building enhancers from the ground up: A synthetic biology approach. *Cell*, 146(1), pp.105–118.

Avican, K. et al., 2015. Reprogramming of Yersinia from Virulent to Persistent Mode Revealed by Complex In Vivo RNA-seq Analysis. *PLoS Pathogens*, 11(1), pp.1–28.

Bentzmann, S. De, Aurouze, M. & Filloux, A., 2006. FppA , a Novel *Pseudomonas aeruginosa* Prepilin Peptidase Involved in Assembly of Type IVb Pili. , 188(13), pp.4851–4860.

Bernard, C.S. et al., 2009. Organization and PprB-Dependent Control of the *Pseudomonas aeruginosa tad* Locus, Involved in Flp Pilus Biology. *Journal of bacteriology*, 191(6), pp.1961–1973.

Boucher, J.C., Schurr, M.J. & Deretic, V., 2000. Dual regulation of mucoidy in *Pseudomonas aeruginosa* and sigma factor antagonism. *Molecular Microbiology*, 36(2), pp.341–351.

Buck, M. et al., 2000. The Bacterial Enhancer-Dependent ς54(ςN) Transcription Factor. *Journal of bacteriology*, 54(15), pp.4129–4136.

Burgess, R.R. & Travers, A.A., 1969. Factor Stimulating Transcription by RNA Polymerase. *Nature*, 221.

D'Argenio, D.A. et al., 2007. Growth phenotypes of *Pseudomonas aeruginosa* lasR mutants adapted to the airways of cystic fibrosis patients. *Molecular Microbiology*, 64(2), pp.512–533.

Damkiaer, S. et al., 2013. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proceedings of the National Academy of Sciences*, 110(19), pp.7766–7771.

Dasgupta, N. et al., 2003. A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Molecular Microbiology*, 50(3), pp.809–824.

Dong, T.G. & Mekalanos, J.J., 2012. Characterization of the RpoN regulon reveals differential regulation of T6SS and new flagellar operons in *Vibrio cholerae* O37 strain V52. *Nucleic acids research*, 40(16), pp.7766–75.

Flores, A.R. et al., 2015. A Single Amino Acid Replacement in the Sensor Kinase LiaS Contributes to a Carrier Phenotype in Group A Streptococcus. *Infection and Immunity*, 83, pp.4237–4246.

Green, M.R. & Sambrook, J., 2012. *Molecular Cloning - A Laboratory Manual*,

Gruber, T.M. & Gross, C. a, 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual review of microbiology*, 57, pp.441–466.

Han, K. et al., 2013. Extraordinary expansion of a *Sorangium cellulosum* genome from an alkaline

milieu. *Scientific reports*, 3, p.2101.

Horstmann, N. et al., 2011. Distinct single amino acid replacements in the control of virulence regulator protein differentially impact streptococcal pathogenesis. *PLoS Pathogens*, 7(10).

Lieberman, T.D. et al., 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics*, 43(12), pp.1275–1280.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), pp.1–34.

Lunter, G. & Goodson, M., 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, 21(6), pp.936–939.

Marvig, R.L. et al., 2013. Genome Analysis of a Transmissible Lineage of *Pseudomonas aeruginosa* Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators. *PLoS Genetics*, 9(9).

Osborne, S.E. et al., 2009. Pathogenic adaptation of intracellular bacteria by rewiring a cis-regulatory input function. *Proceedings of the National Academy of Sciences of the United States of America*, 106(10), pp.3982–7.

Österberg, S., Peso-Santos, T. Del & Shingler, V., 2011. Regulation of Alternative Sigma Factor Use. *Annual Review of Microbiology*, 65(1), pp.37–55.

Ouyang, Z. et al., 2012. Activation of the RpoN-RpoS regulatory pathway during the enzootic life cycle of Borrelia burgdorferi. *BMC microbiology*, 12(1), p.44.

Ouyang, Z., Deka, R.K. & Norgard, M. V., 2011. BosR (BB0647) controls the RpoN-RpoS regulatory pathway and virulence expression in *Borrelia burgdorferi* by a novel DNA-binding mechanism. *PLoS Pathogens*, 7(2).

Potvin, E., Sanschagrin, F. & Levesque, R.C., 2008. Sigma factors in *Pseudomonas aeruginosa*. *FEMS Microbiology Reviews*, 32(1), pp.38–55.

Rustad, T.R. et al., 2014. Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biology*, 15(11), p.502.

Salgado, H. et al., 2013. RegulonDB v8.0: Omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research*, 41(D1), pp.203–213.

Schulz, S. et al., 2015. Elucidation of Sigma Factor-Associated Networks in *Pseudomonas aeruginosa* Reveals a Modular Architecture with Limited and Function-Specific Crosstalk. *PLOS Pathogens*, 11(3), p.e1004744.

Shankaranarayanan, P. et al., 2012. Single-tube linear DNA amplification for genome-wide studies using a few thousand cells. *Nature protocols*, 7(2), pp.328–38.

Shea, P.R. et al., 2011. Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proceedings of the National Academy of*

Sciences of the United States of America, 108(12), pp.5039–44.

Smith, A.H. et al., 2007. Evidence that RpoS (Sigma S) in *Borrelia burgdorferi* is controlled directly by RpoN (Sigma 54/Sigma N). *Journal of Bacteriology*, 189(5), pp.2139–2144.

Smith, E.E. et al., 2006. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc.Natl.Acad.Sci.U.S.A*, 103(22), pp.8487–8492.

Srikumar, S. et al., 2015. RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium. *PLoS Pathogens*, 11(11), pp.1–26.

Stover, C.K. et al., 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406(6799), pp.959–964.

Taylor, M. et al., 1996. The RpoN-box motif of the RNA polymerase sigma factor sigma N plays a role in promoter recognition. *Molecular microbiology*, 22(5), pp.1045–54.

Taylor, T.B., Mulley, G., Dills, A.H., et al., 2015. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science*, 347(6225), pp.1014–1018.

Taylor, T.B., Mulley, G., Mcguffin, L.J., et al., 2015. Evolutionary rewiring of bacterial regulatory networks. *Microbial cell*, 2(7), pp.256–258.

Toledo-Arana, A. et al., 2009. The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, 459(7249), pp.950–956.

Tomich, M., Planet, P.J. & Figurski, D.H., 2007. The *tad* locus : postcards from the widespread colonization island. *Nature Reviews Microbiology*, 5(May), pp.363–375.

Valentini, M., Storelli, N. & Lapouge, K., 2011. Identification of C(4)-dicarboxylate transport systems in *Pseudomonas aeruginosa* PAO1. *Journal of bacteriology*, 193(17), pp.4307–16.

Winsor, G.L. et al., 2016. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic acids research*, 44(November 2015), pp.646–653.

Yang, L. et al., 2011. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), pp.7481–6.

Zhang, Y. et al., 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9).

Figures

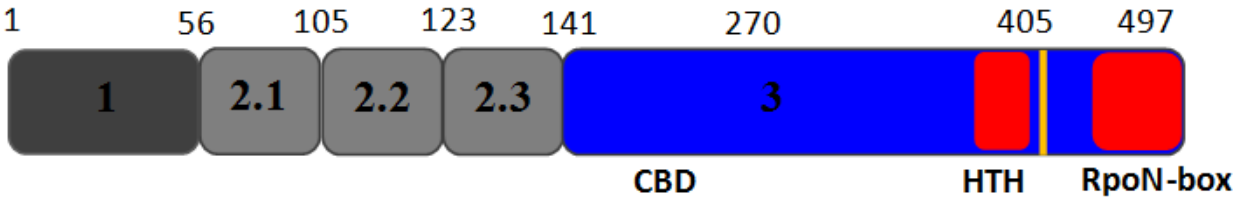


Figure 1. Sequence representation of *P. aeruginosa* RpoN. Grey; *P. aeruginosa* regions 1 and 2, blue; *P. aeruginosa* region 3, red; *P. aeruginosa* HTH and RpoN-box. CBD denotes the core binding domain. The $RpoN^{DK2}$ mutation is marked yellow. Numbers above figure indicate the amino acid sequence number, determined by sequence alignment to the *E. coli* RpoN protein.

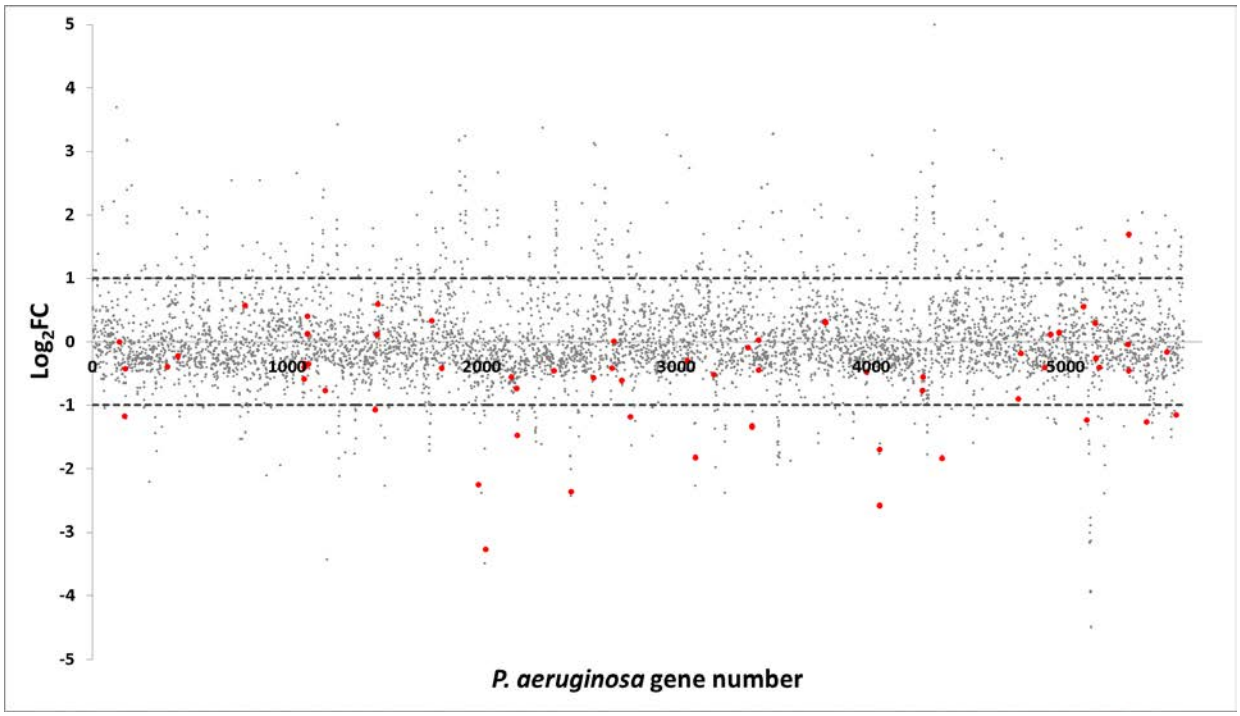


Figure 2. Expression profile of $PAO1^{rpoN(DK2)}$ expressing $RpoN^{DK2}$ compared to $PAO1^{rpoN(WT)}$ expressing $RpoN^{WT}$ (grey series). Genes that are regulated directly by RpoN, i.e. genes whose promoters are known to directly bind to RpoN are marked red. Dotted lines indicate the Log_2FC cut off.

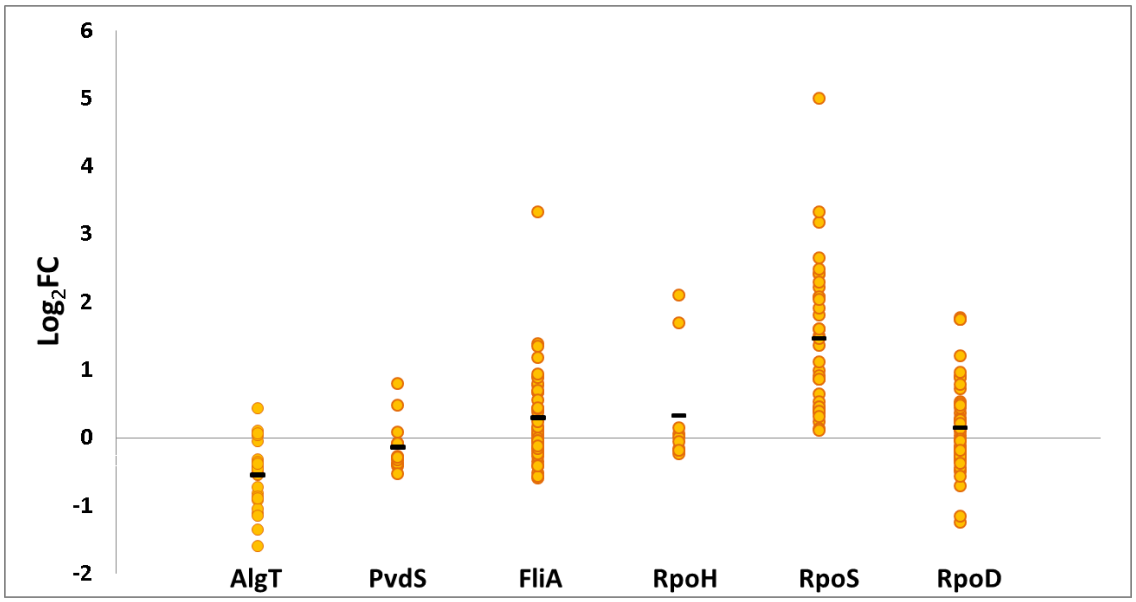


Figure 3. Gene expression values of PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} compared to PAO1$^{rpoN(WT)}$ expressing RpoN^{WT} mapped to known sigma factor core regulons (Schulz et al. 2015). Yellow series denote the specific expression values of each gene in the corresponding regulon, while black bars denote the average Log_2FC for each regulon.

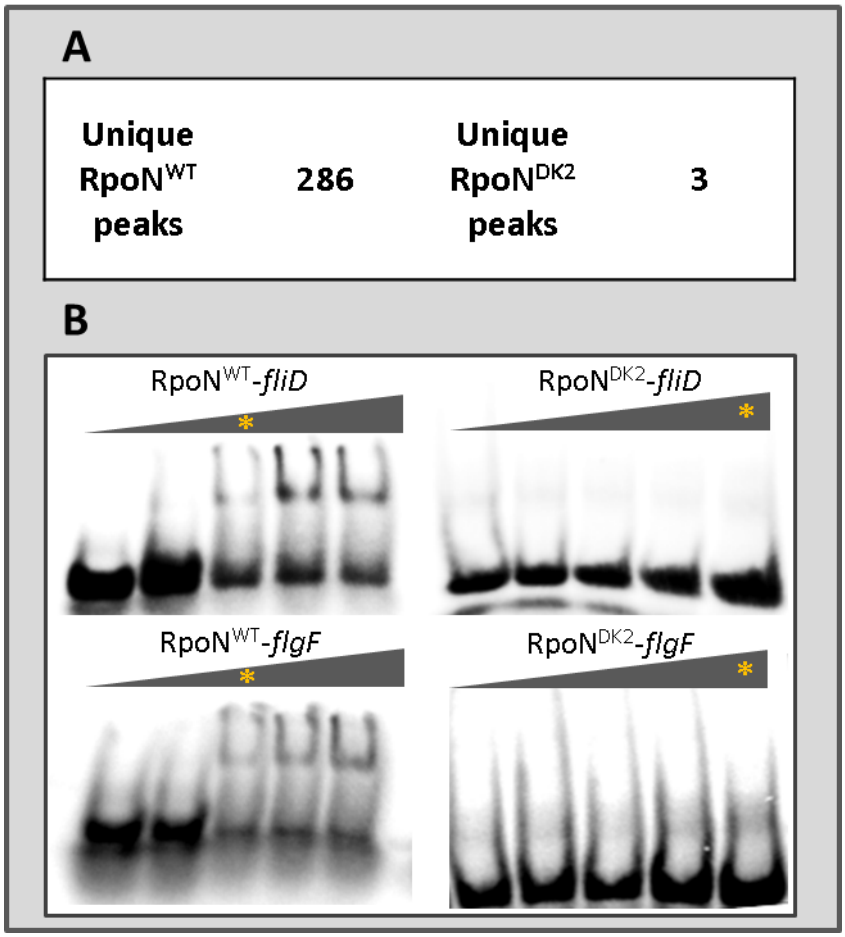


Figure 4. (A) Genome wide binding site distribution of a PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} compared to a PAO1$^{rpoN(WT)}$ expressing RpoN^{WT}. (B) EMSA profile of the interactions of RpoN^{WT} and RpoN^{DK2} to the *fliD* and *flgF* promoter. (B, upper-left) RpoN^{WT} binding to the *fliD* promoter with increasing concentrations of RpoN^{WT} (0 µM, 0.5 µM, 1 µM, 1.5 µM, 2.1 µM). (B, upper-right) RpoN^{DK2} binding to the *fliD* promoter with increasing concentrations of RpoN^{DK2} (0 µM, 0.3 µM, 0.5 µM, 0.8 µM, 1.1 µM). (B, lower-left) RpoN^{WT} binding to the *flgF* promoter with increasing concentrations of RpoN^{WT} (0 µM, 0.5 µM, 1 µM, 1.5 µM, 2.1 µM). (B, lower-right) RpoN^{DK2} binding to the *flgF* promoter with increasing concentrations of RpoN^{DK2} (0 µM, 0.3 µM, 0.5 µM, 0.8 µM, 1.1 µM). Binding profiles at comparable protein concentrations are marked by an asterisk.

A

PAO1 ID	Gene	ChIP Log_2FC	mRNA Log_2FC
PA4305	*rcpC*	3.36	3.33
PA4306	*flp*	3.36	4.99
PA4307	*pctC*	3.36	0.29

B

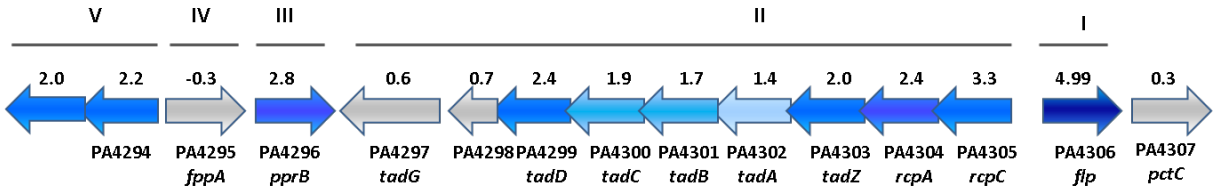
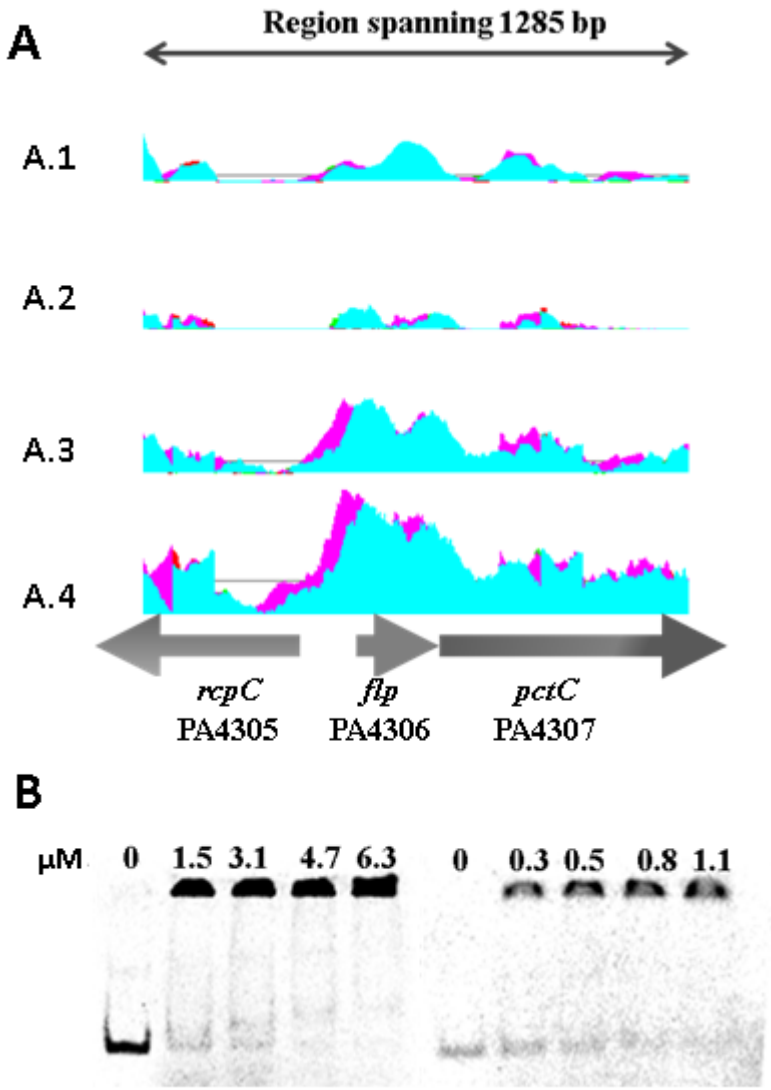


Figure 6. (A) ChIP enrichment and gene expression values for each of the three genes with unique *in vivo* binding to the $RpoN^{DK2}$. The ChIP Log_2FC is reported as a replicate average. (B) Gene expression values of the *flp-tad* locus in a $PAO1^{rpoN(DK2)}$ expressing $RpoN^{DK2}$ compared to a $PAO1^{rpoN(WT)}$ expressing $RpoN^{WT}$. Numbers above gene organisation chart (modified from the *Pseudomonas* Database (Winsor et al. 2016)) indicates the differential expression values, while the top bars and roman numbers indicate organisation of 5 transcriptional units that constitute the *tad* locus (Bernard et al. 2009).



519

520 **Figure 7.** ChIP profile and EMSA analysis demonstrating the remodelled molecular functions of
521 the RpoN^{DK2} protein. (A) A 1285 bp section (genomic coordinates 4829562 – 4830847) of the ChIP
522 profile of the PAO1$^{rpoN(WT)}$ expressing RpoN^{WT} (replicates, A.1 and A.2), as well as the ChIP profile
523 of the PAO1$^{rpoN(DK2)}$ expressing RpoN^{DK2} (replicates, A.3 and A.4). The orientation of the genes
524 *rcpC*, *flp*, and *pctC* are illustrated according to the ChIP profile mapping. (B) EMSA displaying the
525 *in vitro* binding ability of RpoN^{WT} (left) and RpoN^{DK2} (right) to a 305 bp intergenic region between
526 *rcpC* and *flp*.

527

Supplementary figures and tables

Name	Sequence(5´- 3´)	Comments
rpoN_F	AAGAGCCATATGAAACCATCGCTAGTCCTCAAG	Used for amplification of the *rpoN* genes
rpoN_FR	TTCACGAAGCTTCACACCAGTCGCTTGCGCTCG	Used for amplification of the *rpoN* genes
Intreg_F	ACATTGGGGTTATCGACTGG	Used for amplification of intergenic region for use in EMSA
Intreg_R	TTTGCAATACACGAACAGGG	Used for amplification of intergenic region for use in EMSA
FliD promoter	CGGGTTGAACGACTTGGCATGGTGCTTGCCCTATCGA AGGGATA	Used for EMSA. Synthesised as duplex oligo.
FlgF promoter	GTTTTTTCGAATTCTGGCACGGCGCTTGCTGGATAACC TGCAAG	Used for EMSA. Synthesised as duplex oligo.

Table S1 – Sequences for primers and duplex oligos.

PAO1 identifier	DK2 identifier	Gene name	PAO1 gene length (DK2 gene length)	Number of amino acid substitutions or deletions	Protein function
PA4293	PADK2_22370	*pprA*	922 (923)	10	Two-component sensor PprA
PA4294	PADK2_22375		168 (154)	15	Hypothetical protein
PA4295	PADK2_22380	*fppA*	160	2	Flp prepilin peptidase A, FppA
PA4296	PADK2_22385	*pprB*	275	3	Two-component response regulator, PprB
PA4297	PADK2_22390	*tadG*	556	2	TadG
PA4298	PADK2_22395		94	1	Hypothetical protein
PA4299	PADK2_22400	*tadD*	245	1	TadD
PA4300	PADK2_22405	*tadC*	303	0	TadC (pilC homologue)
PA4301	PADK2_22410	*tadB*	294	3	TadB (pilC homologue)
PA4302	PADK2_22415	*tadA*	421	0	TadA ATPase
PA4303	PADK2_22420	*tadZ*	394	0	TadZ
PA4304	PADK2_22425	*rcpA*	416	1	RcpA
PA4305	PADK2_22430	*rcpC*	303	2	RcpC
PA4306	PADK2_22435	*flp*	72	2	Type IVb pilin, Flp

Table S2 –Sequence alignments of the genes constituting the *tad* locus from *P. aeruginosa* PAO1 and *P. aeruginosa* DK2

Epistasis and sigma factor competition in *Pseudomonas aeruginosa*

Eva Kammer Andresen[1], Denitsa Eckweiler[2], Sebastian Schulz[2], Grith Mirriam Maigaard Hermansen[1], Susanne Haussler[2], Maher Abou Hachem[1], Lars Jelsbak[1*]

[1] Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[2] Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany

*Corresponding author: DTU Systems Biology, Building 301, DK2800, Denmark, lj@bio.dtu.dk, +4545256129

Keywords: *Pseudomonas aeruginosa*, evolution, transcriptional regulation, sigma factor competition, epistasis, surface plasmon resonance, ChIP-seq, mRNA seq

Abstract

Evolution of transcriptional regulatory networks is a complex and not well understood process which often involves epistatic interactions between genes encoding regulatory proteins. While a growing body of work have analysed global regulator proteins and how one mutation may affect the regulatory response, our knowledge of the specific molecular mechanisms as well as perturbations in network structure from the overall sum of mutations is very limited. Here, we investigate the combined transcriptional regulatory network dynamics resulting from mutations in two sigma factor proteins in *Pseudomonas aeruginosa*. We use Surface Plasmon Resonance (SPR) to study the direct molecular effects of each mutation, as well as gene expression profiling and Chromatin Immunoprecipitation (ChIP-seq) to study how each mutation alters their respective regulatory network, as well as the epistatic effects produced by the combination of mutations. We show that gene regulatory networks are evolvable structures that may be remodelled either by a switch in sigma factor competition, caused by a direct decrease in the binding affinity to the core RNAP, or they may be altered on a more complex scale, involving a specificity factor, presumably ppGpp, to mediated alteration in gene expression and sigma factor competition. We show that epistasis emerge as a result of global regulator remodelling, and that this effect is the direct cause of an adapted phenotype. This study shows that sigma factor competition and epistasis is closely connected, and that remodelling of sigma factor competition is yet another tool bacteria use to adapt to new environments.

Introduction

Bacteria are faced with continuously changing environments which require the cells to constantly evolve and adapt to accommodate the need for new phenotypes. Mutations in global regulator proteins alter and rewire the cells gene regulatory potential to open up new gene expression routes which enables the bacterium to survive in the constant changing environment. Regulators of gene regulatory networks are often the target for adaptive mutations (Yang et al. 2011; Hindré et al. 2012) and due to their central role in gene regulation, any alteration of their function may produce large downstream impact. Predicting the genotype-phenotype correlation is therefore a central challenge of evolutionary biology, and a number of recent publications have shown that global regulator mutations often result in unexpected pleiotropic phenotypes besides providing the bacterium with an adaptive advantage. For example, modulation of a sensor kinase component of a two-component signal transduction system caused by a 7 bp frameshift caused extensive remodeling of the transcriptomic profile with a resulting phenotypic variation and disease specificity in the Group A Streptococci (Sumby et al. 2006). Other mutations in global regulators have found to dramatically alter a microorganisms host tropism (Viana et al. 2015) or the ability to adapt to complex environments through metabolic selection (Saxer et al. 2014). Characteristic for these studies is that even with the underlining of the importance of single global regulator mutations, the exact molecular mechanisms that facilitate these adapted phenotypes are largely unknown.

While one global regulator mutation may cause pleiotropic effects on both regulatory network organisation and phenotypic outcomes, it is known that variations in the cells genetic composition may cause substantial alterations in the transcription profile of a single global gene regulator (Chugani et al. 2012). As evolution is not a one-step process, and large variation may exist even within the genomic composition of the same bacterial strain, the epistatic effects on the structure of

regulatory network are an important aspect of the evolution of regulatory networks. Epistatic interactions are, however, still poorly understood at both the molecular level as well as how they affect entire gene regulatory networks and thus, and thus, they are consequently difficult to predict.

The evolutionary dynamics of the transmissible *P. aeruginosa* DK2 lineage has been studied extensively in relation to its adaptation to the cystic fibrosis (CF) lung environment, and this system represents a well-established model system for investigations of microbial evolutionary dynamics within a natural environment. During adaptation to the CF lung environment, mutations in regulators of several large gene regulatory networks cause extensive remodelling of gene expression profiles and phenotypes, allowing *P. aeruginosa* to evade both host immune defences and aggressive antibiotic treatment (Yang et al. 2011; Damkiaer et al. 2013). One of the most extensive characteristic adapted phenotypes of *P. aeruginosa* is the conversion to a mucoid phenotype due to a constitutive production of the extracellular polysaccharide alginate and known as a hallmark of chronic infection and a poor prognosis for the patients (Folkesson et al. 2012).

The alternative sigma factor AlgT and its anti-sigma factor MucA directly controls expression of the alginate biosynthetic genes, and mutations in these two genes are well known to affect alginate production and mucoid conversion (Xie et al. 1996). However, genetic determinants controlling the mucoid phenotype involves not only remodelling of the AlgT-MucA gene regulatory network, but may involve other large regulatory networks, such as those controlled by the sigma factor RpoN (Boucher et al. 2000), the transcriptional regulator LasR or even the housekeeping sigma factor RpoD (Damkiaer et al. 2013). Alginate overproduction is thus a specific adapted phenotype, caused by extensive remodelling and epistatic effects from several gene regulatory networks.

The interplay between AlgT and RpoD is especially interesting, as we have previously observed a mucoid phenotype which is dependent on a specific sequential combination of regulatory mutations.

Initially, inactivation of the anti-sigma factor MucA releases AlgT, which then associates with the core RNAP and activates alginate biosynthesis by binding and transcribing from the *algD* promoter, resulting in a mucoid phenotype. Subsequent, a one amino acid substitution in AlgT reverts the phenotype back to nonmucoid, followed by a one amino acid deletion in the housekeeping sigma factor RpoD, which again give rise to a mucoid phenotype (PAO1$^{\Delta mucA(DK2),algT(DK2),rpoD(DK2)}$). The sequential combination of regulatory mutations and the shift between the mucoid and nonmucoid phenotypes implies that constitutive alginate production in PAO1$^{\Delta mucA(DK2),algT(DK2),rpoD(DK2)}$ result from a complex epistatic interplay between the specific mutations in global regulators. In order to fully understand how epistatic interactions between two seemingly unrelated mutations in global regulators can give rise to this phenotype, we need a comprehensive understanding of the molecular basis behind each of the involved regulators, as well as an understanding of the molecular interaction behind the epistatic interactions.

Here, we investigate the gene regulatory network dynamics and molecular mechanisms of mutations in two global regulators, AlgT and RpoD, known to cause an adaptive advantage for *P. aeruginosa* in the form of a mucoid phenotype. We investigate the molecular consequences of each sigma factor mutation, as well as the epistatic effects on the network dynamics. By using gene expression profiling, ChIP-seq, as well as *in vitro* protein-protein interaction assays, we show that gene regulatory network dynamics are regulated on a small scale by adjusting simple mechanisms such as the affinity between a sigma factor and the RNA core polymerase. We also show that, for the evolved PAO1$^{\Delta mucA(DK2),algT(DK2),rpoD(DK2)}$ phenotype, mucoidy is the direct result of epistatic effects, and not a result of alteration of the sigma function factor *per se*, and we suggest the involvement of ppGpp as the effector molecule causing these epistatic effects.

This study links the direct, functional effects of two naturally occurring global regulator mutations to the resulting network displacement, epistatic effects and resulting phenotype, and our results clearly demonstrate the need for achieving a deeper understanding of the evolution of these complex networks.

Materials and Methods

Genetic manipulations

DNA extraction, treatment with modification enzymes and restriction endonucleases, ligation of DNA fragments, and transformation of *Eschericia coli* were performed using standard methods (Green & Sambrook 2012).

Strains and plasmids

All strains and vectors are listed in Table 1, and primers are listed in Table 2.

Expression vectors pJN105-$algT^{(WT)}$, pJN105-$rpoD^{(WT)}$ and pJN105-$rpoN^{(WT)}$ were obtained from (Schulz et al. 2015). pJN105-$algT^{(DK2)}$ and pJN105-$rpoD^{(DK2)}$ were constructed by performing site directed mutagenesis on pJN105-$algT^{(WT)}$ and pJN105-$rpoD^{(WT)}$ using the Quick Change Lightning Site Directed Mutagenesis kit (Stratagene) with the primers ChIP-AlgT-F and ChIP-AlgT-R, and ChIP-RpoD-F and ChIP-RpoD-R, respectively. pJN105-$rpoS^{(WT)}$ was constructed by PCR amplifying the *rpoS* gene from DNA prepared from *P. aeruginosa* PAO1 with primers rpoSF and rpoSR, and ligated into pJN105 using restriction sites EcoRI and XbaI. All transformations of *P. aeruginosa* strains were performed by the method of Choi et al. 2006.

Vectors used for the expression of recombinant sigma factors were constructed by PCR amplifying genes encoding $algT^{WT}$ and $rpoD^{WT}$ from genomic DNA prepared from *P. aeruginosa* PAO1, and PCR amplifying genes encoding $algT^{DK2}$ and $rpoD^{DK2}$ from genomic DNA prepared from PAO1$^{\Delta mucA,algT(DK2)}$ and PAO1$^{rpoD(DK)}$, respectively. $algT^{WT}$ and $algT^{DK2}$ were amplified with the

primers AlgT_F and AlgT_R, and *rpoD*^{WT} and *rpoD*^{DK2} were amplified with the primers RpoD_F and RpoD_R. Genes were ligated into plasmid pET28 (Novagen, Merck Millipore) using restriction sites NdeI and HindIII, producing N-terminal 6xHistidine tagged sequences of each insert. All plasmids were verified by Sanger sequencing and maintained in *E. coli* DH5α.

Expression and Purification of recombinant proteins

The recombinant vectors generated above were transformed into the expression strain *E. coli* Rosetta (DE3) (Novagen, Merck Millipore). For expression of pET28-*algT*$^{(WT)}$, pET28-*algT*$^{(DK2)}$, and pET28-*rpoD*$^{(WT)}$, cells were grown in LB medium supplemented with 10 mM glucose, 50 µg/ml kanamycine, and 34 µg/ml chloramphenicol in 1 L cultures at 30°C to an OD_{600} of 0.5. Expression was induced by adding IPTG to a final concentration of 100 µM and the cultures were harvested after 3 hours and stored at -20°C until use. For expression of pET28-*rpoD*$^{(DK2)}$, cells were grown in LB medium supplemented with 10 mM glucose, 50 µg/ml kanamycine, and 34 µg/ml chloramphenicol in 1 L cultures at 30°C to an OD_{600} of 0.5, then shifting the temperature to 16°C and inducing expression by adding IPTG to a final concentration of 100 µM. Cultures were allowed to express over night, after which cells were harvested and stored at -20°C until use.

Recombinant AlgT^{WT}, AlgT^{DK2} and RpoD^{WT} proteins were all purified following the same procedure. 5g frozen cells were resuspended in 30 ml buffer A (10 mM HEPES pH 7.5, 15 mM Imidazole, 0.5M NaCl and 10% Glycerol), supplemented with 7µl Benzonase (Sigma-Aldrich) and 1 cOmplete protease inhibitor tablet (Roche Diagnostics) and passed through a French Press 3 times at 1,000 bar. Cell lysates were centrifuged and the clarified fraction was loaded onto a 5ml HisTrap column (GE Healthcare) at 0.5 ml/min and 4°C, washed with 6 column volumes (CV) of buffer A, after which bound fractions were eluted with buffer B (10 mM HEPES pH 7.5, 400 mM Imidazole, 0.5 M NaCl and 10% Glycerol) in 30 CV. Fractions containing 90%< pure AlgT^{WT}, AlgT^{DK2} and RpoD^{WT} were collected and dialysed against storage buffer (20 mM HEPES pH 7.5, 150 mM NaCl,

5 mM $MgCl_2$, 0.5 mM DTT and 10% glycerol). Proteins were stored at -80°C in storage buffer and glycerol added to a final concentration of 50%.

Purification of $RpoD^{DK2}$ followed the initial procedure until clarification of cell lysate after French Press. After centrifugation, $RpoD^{DK2}$ was purified from inclusion bodies with a protocol adapted from (Anthony et al. 2003). The pelleted cell lysate was resuspended in 30 ml 50 mM Tris–HCl pH 7.9, 100 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT, 5% glycerol and 1% TritonX-100 pr. harvested 2g expression biomass and pelleted by centrifugation. The inclusion body pellet was washed two times with 50 mM Tris–HCl pH 7.9, 100 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT and 5% glycerol to remove residual TritonX-100, and solubilised in 15 ml 50 mM Tris–HCl pH 7.9, 6 M guanidine–HCl, 100 M NaCl and 0.1 mM DTT pr. harvested 2g expression biomass. Insoluble material was removed by centrifugation at. The supernatant was then purified on a 5 ml HisTrap column, which also facilitated refolding of the $RpoD^{DK2}$. After loading of the sample, the column was washed with 4 CV of 100% buffer A, 5 CV of 5% buffer B, eluted over 15 CV with a 5-50% gradient of buffer B and finished with 3 CV of 100% buffer B. Everything at 0.5 ml/min. All steps of this protocol were performed at 4°C, and all buffers were cooled to at 4°C before use.

Surface Plasmon Resonance analysis of sigma factor-core-RNAP interactions

Surface Plasmon Resonance (SPR) measurements were performed on a Biacore T100 instrument (GE Healthcare) with Sensor Chip NTA (GE Healtcare). Purified, 6xHistidine tagged sigma factor was used as ligand and immobilised on the chip by initially following suppliers recommendations for stripping and activating the chip, after which sigma factor was applied in running buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM DTT and 10% glycerol) to channel 2 until 2-300 Response Units (RU) was reached. For analysis of sigma factor – core RNAP interactions, *E. coli* core RNAP (New England Biolabs) was injected with a flowrate of 30 µl/min into both channels for 180 seconds, followed by a 400 second dissociation time. Regeneration was performed

by injecting 2.5 mM NaOH at 30 µl/min for 15 seconds in both channels. All measurements were performed in duplicates at 15°C.

Interactions were analysed using the Biacore T100 Evaluation Software (Biacore, GE Healthcare). For each analysis, the response of the reference channel was subtracted from the interaction. Replicates were then fitted to a 1:1 (Langmuir) interaction model. Equilibrium binding constants were calculated from the ratio K_{off} /K_{on} over a range of protein concentrations.

In vivo overexpression of sigma factor levels

Overexpression of each sigma factor was performed by growing strains for 8 hours in LB medium supplemented with 30 µg/ml gentamycin to maintain constructs and 1% glucose to supress leaky expression. After 8 hours, cultures were plated on LB-agar plates containing either 30 µg/ml gentamycin and 1 % glucose as a non-induced control, or 30 µg/ml gentamycin and 0.3% arabinose to induce sigma factor expression. Plates were incubated overnight, and colony morphologies were recorded using a CoolSNAPPro cf color camera mounted on a Zeiss Axioplan2 microscope.

Chromatin-Immunoprecipitation followed by deep sequencing

Chromatin immunoprecipitation was performed on PAO1$^{\Delta mucA,algT(WT)}$ expressing pJN105-*algT*$^{(WT)}$ and PAO1$^{\Delta mucA,algT(DK2)}$ expressing pJN105-*algT*$^{(DK2)}$ to compare the genome wide binding profiles of AlgT^{WT} and AlgT^{DK2}, and on PAO1 expressing pJN105-*rpoD*$^{(WT)}$ and PAO1$^{rpoD(DK2)}$ expressing pJN105-*rpoD*$^{(DK2)}$ to compare the genome wide binding profiles of RpoD^{WT} and RpoD^{DK2}.

ChIP experiments were performed as described previously (Schulz et al. 2015). Briefly, 2x50 ml culture was grown to an OD_{600} of 2.0 for PAO1$^{\Delta mucA,algT(WT)}$ expressing pJN105-*algT*$^{(WT)}$ and PAO1$^{\Delta mucA,algT(DK2)}$ expressing pJN105-*algT*$^{(DK2)}$. PAO1 expressing pJN105-*rpoD*$^{(WT)}$ and PAO1$^{rpoD(DK2)}$ expressing pJN105-*rpoD*$^{(DK2)}$ were grown to an OD_{600} of 0.6. Sigma factor expression was then induced with 0.5% l-arabinose and AlgT^{WT} and AlgT^{DK2} were allowed to express for 75 minutes while RpoD^{WT} and RpoD^{DK2} were allowed to express for 45 minutes. Before

formaldehyde treatment, 2x1.5 ml was sampled for mRNA profiling from each sample. Crosslinking was performed by adding 0.5% formaldehyde and the crosslinking reaction was quenched after 5 minutes by adding glycine to a final concentration of 137 mM. Cells were pelleted and resuspended in 0.5 ml lysis buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 10 mM EDTA, 20% sucrose), and DNA was fragmented to an average size of 200-250 bp using a Covaris E220 Sonicator with the setting 20% duty cycle, Intensity 7, 200 cycles pr. burst for 30 minutes. Immunoprecipitation was performed with 15 µl of anti-6xHistag antibody (ab9108, Abcam) overnight at 4°C and incubated with 30 µl Dynabeads Protein G. Protein-DNA complexes were eluted from the beads in 25 µl elution buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 % (m/v) SDS) at 65°C for 15 minutes. Immunoprecipitated samples were treated with RNAse A and proteinase K, and DNA was purified using the QIAquick PCR purification Kit (Qiagen).

DNA was sequenced using the NEBNext Ultra DNA Library Prep Kit for Illumina, and sequenced on a Mi-Seq, paired end sequencing with a read length of 250 bp.

ChIP-seq data were analysed as described previously (Schulz et al. 2015). For all ChIP samples, model-based analysis of ChIP-seq (MACS) (Zhang et al. 2008) was applied for peak detection using a P value cut-off value of 0.05 and shift size 30 for the peak modelling. Promoters were searched within the window [-500nt, TSS, +100nt] and considered significant when they were detected in both ChIP-seq replicates with an enrichment factor of at least 2 and a P value of less than 0.01.

Gene expression profiling

The 1.5 ml samples collected above was used for isolation of total RNA using the Agencourt RNAClean XP Kit (Beckman Coulter) and enrichment of mRNA as well as library preparation for sequencing was performed using the ScriptSeq Complete Kit (Bacteria). Samples were sequenced as paired end 2x76 bp on an Illumina HiSeq 2000.

223 Sequences were mapped to the PAO1 genome using Stampy (Lunter & Goodson 2011) and
224 differential expression was called with the R package DESeq2 (Love et al. 2014) using each
225 biological duplicate. Differentially expressed genes were identified as those having absolute
226 logarithmic fold change greater than one ($|\log_2 FC|>1$) at significance level of less than 5%.

227

Results

The AlgT^{DK2} mutation decrease the affinity for the core RNAP

To infer about the possible functional effects of the K19L mutation in AlgT^{DK2}, we performed a sequence alignment of the *E. coli* and *P. aeruginosa* AlgT sequences and mapped the AlgT^{DK2} mutation to the N-terminal part of region 2 (**Figure 1A**). Since Region 2 has previously been associated with interaction between the sigma factor and core RNAP (Burgess & Anthony 2001), we hypothesized that the AlgT^{DK2} protein would show altered binding affinity core RNAP.

To examine this hypothesis, we used SPR to characterize the interaction between purified, recombinant AlgT^{WT} and AlgT^{DK2} to *E. coli* core RNAP. We determined the affinity between AlgT^{WT} and core RNAP to be 0.23 nM, while the affinity between AlgT^{DK2} and core RNAP was decreased to 18.1 nM (**Figure 1B** and Figure **S1**.A-B). The consequence of a decreased affinity between AlgT^{DK2} and the core RNAP is a decreased ability to form holo RNAP complexes, and therefore most likely a downregulation of the entire set of the AlgT controlled genes in a PAO1$^{\Delta mucA,algT(DK2)}$ background.

Genome wide binding profile of AlgT is unaffected by the AlgT^{DK2} mutation

We next sought to determine, if the altered interaction between AlgT^{DK2} and the core RNAP results in an altered ability for the AlgT^{DK2}-RNAP to recognise and bind to AlgT controlled promoters. By using ChIP-seq, we were able to compare the genome wide binding site distribution of AlgT^{WT} expressed in PAO1$^{\Delta mucA,algT(WT)}$ to the binding site distribution of AlgT^{DK2} expressed in PAO1$^{\Delta mucA,algT(DK2)}$. Besides 3 unique peaks called for the AlgT^{WT} (Dataset S1), the two binding profiles were identical (**Figure 1C**), indicating that the AlgT^{DK2}-RNAP maintain its ability bind to AlgT controlled promoters.

Due to the important role of AlgT in alginate production, we further inspected the binding profiles at the *algD* promoter region to verify that both $AlgT^{WT}$-RNAP and $AlgT^{DK2}$-RNAP were able to bind to the *algD* promoter. The identical binding profiles confirm that both the $AlgT^{WT}$-RNAP and the $AlgT^{DK2}$ -RNAP holoenzymes are able to bind to the *algD* promoter (**Figure 1D**).

Gene regulatory effects of the $AlgT^{DK2}$ variant

We then examined the transcription profile of strains expressing either $AlgT^{WT}$ or $AlgT^{DK2}$ to determine how the downstream regulatory system was affected by the decreased binding affinity of the $AlgT^{DK2}$ protein to the core RNAP.

At a significance level of $p<0.05$ and log fold change ($|\log_2 FC|>1$, we discovered 788 genes to be upregulated and 1093 genes be to be downregulated in $PAO1^{\Delta mucA,algT(DK2)}$ expressing $AlgT^{(DK2)}$ compared to $PAO1^{\Delta mucA,algT(WT)}$ expressing $AlgT^{(WT)}$ (**Figure 2** and dataset S2). Of these, 98 genes were heavily upregulated, as well as 61 genes were heavily downregulated with ($|\log_2 FC|>3$, with the entire 12 genes of the alginate biosynthesis operon displaying the most critical downregulation, indicating that the decreased binding affinity of the $AlgT^{DK2}$ –RNAP is the direct result of the nonmucoid phenotype associated with the $algT^{DK2}$ mutation.

Sigma factor competition is a direct trigger of phenotype changes in the $PAO1^{\Delta mucA,algT(DK2)}$ variant

We next used the gene expression dataset to investigate how the decreased binding affinity of the $AlgT^{DK2}$ protein to the core RNAP was reflected on the known AlgT regulon. By mapping expression values of $PAO1^{\Delta mucA,algT(DK2)}$ expressing $AlgT^{(DK2)}$ compared to $PAO1^{\Delta mucA,algT(WT)}$ expressing $AlgT^{(WT)}$ to the core regulon of AlgT (defined as genes containing promoters known to interact directly with AlgT) (Schulz et al. 2015), we discovered that the negative effect on expression values was not limited to the *algD* operon and alginate synthesis, but rather that the entire core AlgT regulon was negatively affected by the $AlgT^{DK2}$ mutation (**Figure 3**, blue series).

However, this analysis also revealed an AlgT^{DK2} specific positive effect on the RpoS regulon with >90% of the direct RpoS controlled genes being positively affected by the AlgT$^{(DK2)}$ mutation (**Figure 3A**, yellow series). A similar positive effect was observed for the FliA regulon with 88% of the direct regulon being upregulated as a response to the AlgT$^{(DK2)}$ mutation (data not shown). The specific downregulation of the AlgT regulon and upregulation of the RpoS regulon suggest that the decreased binding affinity between AlgT$^{(DK2)}$ and the core RNAP causes a direct remodeling of sigma factor competition, resulting in downregulation of AlgT controlled genes, including the *algD* operon.

To confirm the involvement of sigma factor competition in the switch from mucoid to a nonmucoid phenotype, we overexpressed AlgT$^{(DK2)}$ in PAO1$^{\Delta mucA,algT(DK2)}$ in an attempt reverse the nonmucoid phenotype of the PAO1$^{\Delta mucA,algT(DK2)}$ to the mucoid phenotype of the PAO1$^{\Delta mucA,algT(WT)}$. Indeed, overexpression of AlgT$^{(DK2)}$ in PAO1$^{\Delta mucA,algT(DK2)}$ did revert the phenotype back to mucoid (**Figure 3B**), and we conclude that this effect is due to restoration of the AlgT^{DK2}-RNAP equilibrium, which then reestablishes transcription from the *algD* promoter.

Our findings demonstrate that a decreased affinity for the core RNAP directly affects sigma factor competition by lowering the ability of AlgT^{DK2} to compete for the core RNAP, and that this effect can be compensated by an increased numbers of AlgT^{DK2} molecules.

RpoD^{DK2} is a functional neutral mutation

To infer about the possible functional effects of the RpoD^{DK2} ΔE507 mutation, we performed a sequence alignment of the *E. coli* and *P. aeruginosa* RpoD sequence and mapped the RpoD^{DK2} mutation to region 3.2 (**Figure 4A**). RpoD consisting of 4 regions, each involved in various functions. Region I is involved in modulation of DNA binding and some inhibition of DNA binding. Region 2 is subdivided into 4 regions each involved in functions including binding to the

core RNAP, melting of DNA and recognition of the -10 promoter element. Region 4 is involved in binding to the -35 promoter region and interaction to transcriptional activators (Paget & Helmann 2003; Potvin et al. 2008; Österberg et al. 2011). Region 3 is divided into two subregions, 3.1 and 3.2, and these regions has been associated with several functions such as the binding of initiating NTPs, RNA priming, and promoter recognition, opening, and escape (Severinov et al. 1994; Kulbachinskiy & Mustaev 2006; Pupov et al. 2014). Specific amino acid substitutions of region 3.2 has also been shown to suppress auxotrophy of an *E. coli* ppGpp° strain, as well as being involved in sigma affinity for the core RNAP (Zhou et al. 1992; Hernandez & Cashel 1995; Cashel et al. 2003) .

We initially investigated whether the $RpoD^{DK2}$ mutation resulted in a decreased binding to the core RNAP, as was the case for $AlgT^{DK2}$, by characterising the interaction of $RpoD^{WT}$ and $RpoD^{DK2}$ to *E. coli* core RNAP. However, we found that the affinities of both $RpoD^{WT}$ and $RpoD^{DK2}$ to the core RNAP were similar (**Figure 4B** and Figure S1.C-D). This indicates that directly altered sigma factor competition is not the cause of the shift from the nonmucoid phenotype of $PAO1^{\Delta mucA,algT(DK2)}$ to the mucoid phenotype of $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$.

We next sought to determine if the molecular effect of the mutation was to be found at the regulatory level, for example by directly affecting the ability of the $RpoD^{DK2}$ holo RNAP to recognise its promoter sequences *in vivo*. We therefore performed ChIP-seq to compare the genome wide binding profile of $RpoD^{DK2}$ expressed in $PAO1^{rpoD(DK2)}$ to the binding profile of $RpoD^{WT}$ expressed in $PAO1^{rpoD(WT)}$. Comparison of the two ChIP profiles showed only few unique binding sites (**Figure 4C**), and compared to the number of known RpoD binding sites in *P. aeruginosa* PA14 (Schulz et al. 2015, 308 $RpoD^{WT}$ binding sites), these sites few most likely holds no biological relevance.

Gene expression profiling of the RpoD^{DK2} variant

As neither affinity to the core RNAP, nor the ability of the holo RNAP to recognise DNA promoter sequences was affected by the RpoD^{DK2} mutation, we investigated the gene expression profile of PAO1$^{rpoD(DK2)}$ expressing RpoD^{DK2} compared to PAO1$^{rpoD(WT)}$ expressing RpoD^{WT} to identify any leads on the molecular effects of the RpoD^{DK2} mutation.

Interestingly, the expression profile showed only very subtle changes in response to the RpoD^{DK2} mutation (**Figure 5**, black series and dataset S4). At a significance level of $p<0.05$ and ($|\log_2 FC|$)>1, we observed only minor impacts with 49 genes being downregulated and 18 genes being upregulated, corresponding to only 1.2% of the total number of genes in *P. aeruginosa*. We did not observe any notable regulation of the genes involved in alginate biosynthesis, indicating that the RpoD^{DK2} mutation affects neither alginate biosynthesis genes, nor the activity of the AlgT sigma factor directly.

RpoD^{DK2} regulatory effects and alginate production is a result of epistatic interactions

To investigate if the effects of the RpoD^{DK2} mutant was contingent on the specific genetic environment of PAO1$^{\Delta mucA,algT(DK2),rpoD(DK2)}$, we examined the transcriptional landscape of PAO1$^{\Delta mucA,algT(DK2),rpoD(DK2)}$ expressing RpoD^{DK2} compared to PAO1$^{rpoD(DK2)}$ expressing RpoD^{DK2}. Introducing the Δ*mucA*, *algT*$^{(DK2)}$ mutations clearly affects the global transcription profile in a drastic manner (**Figure 5**, grey series, and dataset S5). The epistatic interactions of the 3 mutations results in a drastically altered transcriptional landscape with 1353 genes found to be upregulated (FC>1), and 1714 genes found to be downregulated (FC<-1) at a significance level of $p<0.05$. In addition, the entire alginate biosynthesis operon was upregulated only in the presence of the specific combination of PAO1$^{\Delta mucA,algT(DK2),\ rpoD(DK2)}$, indicating that the phenotypic shift from nonmucoid to mucoid is contingent on the specific combination of mutations and the interplay of the AlgT-RpoD regulatory networks.

An AlgT-RpoD regulatory interplay is the main target for regulation of the mucoid phenotype

To investigate the robustness of the $AlgT^{DK2}$ -$RpoD^{DK2}$ mediated mucoid phenotype, we overexpressed the sigma factors RpoN and RpoS, as well as the $RpoD^{WT}$ and $RpoD^{DK2}$ variants in $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$. While we expected that overexpression of RpoN and RpoS would result in a phenotypic shift due to remodelling of sigma factor-holo RNAP distribution, we were not able to switch the mucoid phenotype to nonmucoid (**Figure 6**). Overexpression of RpoN did not affect the colony morphology, while overexpression of RpoS in $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$ produced stressed, small colony variants, which none the less remained mucoid. Only overexpression of $RpoD^{WT}$ and $RpoD^{DK2}$ supported a switch to the nonmucoid phenotype, indicating that the specific regulatory network remodelling leading to the mucoid phenotype target the specific AlgT-RpoD regulatory interplay, but also that this regulatory interplay is robust to other network perturbations.

Discussion

The results presented here demonstrate that gene regulatory networks are subjected to evolutionary modifications that work through different molecular mechanisms. Our aim was obtain a molecular understanding of how mutations in global regulators alter their functions to accommodate new phenotypes and gene expression patterns. Our starting hypothesis which was that the two sigma factor mutations, $AlgT^{DK2}$ and $RpoD^{DK2}$, each resulted in a decreased affinity for the core RNAP, and thus resulted in a specific and isolated remodelling of sigma factor competition was refuted. Instead, we found the specific adapted mucoid phenotype of *P. aeruginosa* is directly dependent on epistatic interactions which assist in rewiring of regulatory networks.

The specific K19E mutation in $AlgT^{DK2}$ reduces affinity for the core RNAP, and directly through this mechanism alters expression of most genes comprising the AlgT regulon. The mutation does

not change the ability of the sigma factor to recognise its promoter sequences, thus creating a specific on-off effect on downstream gene transcription. The involvement of sigma factor competition in the switch between mucoid and nonmucoid phenotypes of *P. aeruginosa* has been addressed previously (Damkiaer et al. 2013; Yin et al. 2013). With our comparison of the actual binding affinities of $AlgT^{WT}$ and $AlgT^{DK2}$ to the core RNAP, we here present direct evidence that mutations may remodel sigma factor competition to accommodate new phenotypes.

The $RpoD^{DK2}$ mutation was discovered due to the sudden emergence of mucoid clinical *P. aeruginosa* CF isolates (Damkiaer et al. 2013). We initially hypothesised that, like $AlgT^{DK2}$, the $RpoD^{DK2}$ mutation would cause a reduced affinity for the core RNAP, thus causing a simple regulatory shift due to reestablishment of sigma factor competition. We found that neither binding to the core RNAP, nor the ability of the $RpoD^{DK2}$-RNAP to recognise promoter sequences was affected by the mutation.

While our studies have not clearly defined the mechanism by which $RpoD^{DK2}$ further rewire the $PAO1^{\Delta mucA,algT(DK2)}$ regulatory network, our gene expression analysis shows that the combination of $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$ creates massive disturbance in the gene regulatory network and activates alginate production, which is not further influenced by activation of the RpoN or RpoS regulon. We speculate that the $RpoD^{DK}$ mutation alters the protein's sensitivity or regulatory response to an unknown factor, which then allows activation the AlgT regulon. A possible candidate for this unknown factor is the stringent response regulator, ppGpp. While ppGpp has not previously been appointed a role in mucoid conversion and adaptation of *P. aeruginosa* to the CF lung environment, its role as a global transcriptional regulator of gene expression during the stringent response is well documented. ppGpp is known to be a negative regulator of rRNA synthesis during starvation, a positive regulator of amino acid biosynthesis and virulence genes, as well as a modulator of sigma factor competition (Magnusson et al. 2005). ppGpp binds to the core RNAP (Ross et al. 2013) and

393 negatively regulates promoters with short linkers between the -35/-10 consensus sequence and GC-
394 rich discriminator sequence, while it results in positive regulation of promoters with longer linkers
395 and AT-rich discriminators (Potrykus & Cashel 2008). When cells experience stress or starvation,
396 binding of ppGpp destabilises the RpoD-RNAP complex at intrinsically unstable rRNA promoters,
397 which may cause an increase in the levels of free core RNAP (Laurie et al. 2003), and thus
398 indirectly modulate sigma factor competition.

399 Interestingly, two mutations in the *E. coli* RpoD (P504L and S506F) have previously been found to
400 be suppressor alleles of a $ppGpp^0$ phenotype which is auxotroph for several amino acids, and it was
401 shown that the RpoD(P504L) and RpoD(S506F) variants displayed either a hypersensitivity to
402 ppGpp, or mimicked the presence of ppGpp, respectively (Hernandez & Cashel 1995).
403 Interestingly, the *P. aeruginosa* $rpoD^{DK2}$ mutation is located in close proximity (Δ507E), which
404 raises the possibility that a similar effect may be the case for the *P. aeruginosa* $RpoD^{DK2}$ mutation.
405 If the $RpoD^{DK2}$ mutation modulates the proteins sensitivity to ppGpp, or mimicks the function of
406 ppGpp, the $RpoD^{DK2}$-RNAP would become destabilised at intrinsically unstable rRNA promoters,
407 which would increase the fraction of free core RNAP available for sigma factor competition. In
408 accordance with this, the direct affinity between the $RpoD^{DK2}$-RNAP would not have to be altered,
409 as shown by our *in vitro* interaction assay. Rather, the increase in the pool of free core would be due
410 instability of the $RpoD^{DK2}$-RNAP at rRNA promoters.

411 From our expression data, however, it is evident that the isolated effect of the $RpoD^{DK2}$ mutation
412 does not cause the observed transcriptional rewiring. What then, could explain the need of both the
413 $AlgT^{DK2}$ and $RpoD^{DK2}$ mutation to produce the specific activation at the algD promoter, observed in
414 the $PAO1^{\Delta mucA,algT(DK2),rpoD(DK2)}$? Among the numerous regulatory functions of ppGpp, it is
415 suggested that ppGpp may also directly alter AlgT activity. *In vitro,* ppGpp and its potentiator
416 DksA directly activate transcription by AlgT-RNAP (Costanzo et al. 2008), and *in vivo*, ppGpp is

required for activation of the AlgT-RNAP under certain starvation conditions (Gopalkrishnan et al. 2014). We speculate that that ppGpp may be involved in both the regulatory effects of $RpoD^{DK2}$, as well as activity of AlgT. ppGpp would then serve as a master regulator of the epistatic effects observed from the combination of the $AlgT^{DK2}$ and $RpoD^{DK2}$ mutations.

Based on these observations, we suggest the following molecular and regulatory model for the stepwise remodelling of the two regulatory networks controlled by AlgT and RpoD; An initial mucoid $PAO1^{\Delta mucA}$ strain is able to produce alginate due to release of AlgT from its anti-sigma factor, MucA that becomes inactivated by mutation. A subsequent mutation in the AlgT sigma factor reduces its affinity for the core RNAP, thereby lowers its ability to compete for holo RNAP formation. The result is abrogation of transcription from the *algD* promoter, and a downregulation of all genes in the core AlgT regulon. Finally, the $RpoD^{DK2}$ mutation causes a modulated sensitivity to ppGpp, which indirectly alters the sigma factor competition landscape and allows $AlgT^{DK2}$ to form holo RNAP, thus initiating transcription from the *algD* promoter.

Though the specific mechanistic involvement of ppGpp remains to be elucidated, our results show that gene regulatory networks may be remodelled directly by changes protein-protein interaction abilities, or they may be remodelled through complex epistatic effects, mediated through direct ppGpp involvement, or by ppGpp mimicry.

Interestingly, any ppGpp mediated rewiring of the two regulatory networks studied here was contingent on the presence of both mutations. We speculate that AlgT and RpoD may be mutational hotspots for regulatory protein modifications to propel the mucoid phenotype, and an interesting starting point for future studies could be whether this regulatory network rewiring presents the only regulatory road to the mucoid phenotype, or if alternative regulatory modifications will create the same effect, through different routes.

Acknowledgements

We would like to thank Susanne Kofoed and Tanja Nicolai for expert technical assistance, and Agata Bielecka for assistance on ChIP- and mRNA experiments. We would also like to thank the Augustinus Foundation and Oticon Foundation for travel grants, as well as the Villum Foundation for funding for this study to LJ (Grant number VKR023113). LJ acknowledges additional funding from the Novo Nordisk Foundation and the Lundbeck Foundation.

Competing interests

None

References

Anthony, L.C. et al., 2003. Expression, purification of, and monoclonal antibodies to sigma factors from *Escherichia coli*. *Methods in enzymology*, 370(1991), pp.181–92.

Boucher, J.C., Schurr, M.J. & Deretic, V., 2000. Dual regulation of mucoidy in *Pseudomonas aeruginosa* and sigma factor antagonism. *Molecular Microbiology*, 36(2), pp.341–351.

Burgess, R.R. & Anthony, L., 2001. How sigma docks to RNA polymerase and what sigma does. *Current opinion in microbiology*, 4(2), pp.126–31.

Cashel, M., Hsu, L.M. & Hernandez, V.J., 2003. Changes in conserved region 3 of *Escherichia coli* σ70 reduce abortive transcription and enhance promoter escape. *Journal of Biological Chemistry*, 278(8), pp.5539–5547.

Choi, K.H., Kumar, A. & Schweizer, H.P., 2006. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. *Journal of Microbiological Methods*,

462 64(3), pp.391–397.

463 Chugani, S. et al., 2012. Strain-dependent diversity in the *Pseudomonas aeruginosa* quorum-
464 sensing regulon. *Proceedings of the National Academy of Sciences of the United States of*
465 *America*, 109(41), pp.E2823–31.

466 Costanzo, A. et al., 2008. ppGpp and DksA likely regulate the activity of the extracytoplasmic
467 stress factor SigmaE in *Escherichia coli* by both direct and indirect mechanisms. *Molecular*
468 *Microbiology*, 67(3), pp.619–632.

469 Damkiaer, S. et al., 2013. Evolutionary remodeling of global regulatory networks during long-term
470 bacterial adaptation to human hosts. *Proceedings of the National Academy of Sciences*,
471 110(19), pp.7766–7771.

472 Folkesson, A. et al., 2012. Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an
473 evolutionary perspective. *Nature Reviews Microbiology*, 10(12), pp.841–51.

474 Gopalkrishnan, S., Nicoloff, H. & Ades, S.E., 2014. Co-ordinated regulation of the
475 extracytoplasmic stress factor, sigmaE, with other *Escherichia coli* sigma factors by (p)ppGpp
476 and DksA may be achieved by specific regulation of individual holoenzymes. *Molecular*
477 *Microbiology*, 93(3), pp.479–493.

478 Green, M.R. & Sambrook, J., 2012. *Molecular Cloning - A Laboratory Manual*,

479 Hernandez, V.J. & Cashel, M., 1995. Changes in conserved region 3 of *Escherichia coli* sigma 70
480 mediate ppGpp-dependent functions *in vivo*. *Journal of molecular biology*, 252(5), pp.536–49.

481 Hindré, T. et al., 2012. New insights into bacterial adaptation through *in vivo* and *in silico*
482 experimental evolution. *Nature Reviews Microbiology*, 10(5), pp.352–365.

Kulbachinskiy, A. & Mustaev, A., 2006. Region 3.2 of the σ subunit contributes to the binding of the 3´-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *Journal of Biological Chemistry*, 281(27), pp.18273–18276.

Laurie, A.D. et al., 2003. The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. *The Journal of biological chemistry*, 278(3), pp.1494–503.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), pp.1–34.

Lunter, G. & Goodson, M., 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, 21(6), pp.936–939.

Magnusson, L.U., Farewell, A. & Nyström, T., 2005. ppGpp: A global regulator in *Escherichia coli*. *Trends in Microbiology*, 13(5), pp.236–242.

Österberg, S., Peso-Santos, T. Del & Shingler, V., 2011. Regulation of Alternative Sigma Factor Use. *Annual Review of Microbiology*, 65(1), pp.37–55.

Paget, M.S.B. & Helmann, J.D., 2003. The Sigma 70 family of sigma factors. *Genome Biology*, pp.1–6.

Potrykus, K. & Cashel, M., 2008. (p)ppGpp: Still Magical? *. *Annual Review of Microbiology*, 62(1), pp.35–51.

Potvin, E., Sanschagrin, F. & Levesque, R.C., 2008. Sigma factors in *Pseudomonas aeruginosa*. *FEMS Microbiology Reviews*, 32(1), pp.38–55.

Pupov, D. et al., 2014. Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic acids research*, 42(7), pp.4494–504.

504 Ross, W. et al., 2013. The magic spot: A ppGpp binding site on *E. coli* RNA polymerase
505 responsible for regulation of transcription initiation. *Molecular Cell*, 50(3), pp.420–429.

506 Saxer, G. et al., 2014. Mutations in Global Regulators Lead to Metabolic Selection during
507 Adaptation to Complex Environments. *PLoS Genetics*, 10(12), p.e1004872.

508 Schulz, S. et al., 2015. Elucidation of Sigma Factor-Associated Networks in *Pseudomonas*
509 *aeruginosa* Reveals a Modular Architecture with Limited and Function-Specific Crosstalk.
510 *PLOS Pathogens*, 11(3), p.e1004744.

511 Severinov, K. et al., 1994. The sigma subunit conserved region 3 is part of "5'-face" of active center
512 of *Escherichia coli* RNA polymerase. *The Journal of biological chemistry.*, 269(33),
513 pp.20826–20828.

514 Sumby, P. et al., 2006. Genome-wide analysis of group A streptococci reveals a mutation that
515 modulates global phenotype and disease specificity. *PLoS Pathogens*, 2(1), pp.0041–0049.

516 Viana, D. et al., 2015. A single natural nucleotide mutation alters bacterial pathogen host tropism.
517 *Nature Genetics*, 47(4), pp.361–366.

518 Xie, Z.D. et al., 1996. Sigma factor-anti-sigma factor interaction in alginate synthesis: inhibition of
519 AlgT by MucA. *Journal of bacteriology*, 178(16), pp.4990–4996.

520 Yang, L. et al., 2011. Evolutionary dynamics of bacteria in a human host environment. *Proceedings*
521 *of the National Academy of Sciences of the United States of America*, 108(18), pp.7481–6.

522 Yin, Y. et al., 2013. Evidence for sigma factor competition in the regulation of alginate production
523 by *Pseudomonas aeruginosa*. *PloS one*, 8(8), p.e72329.

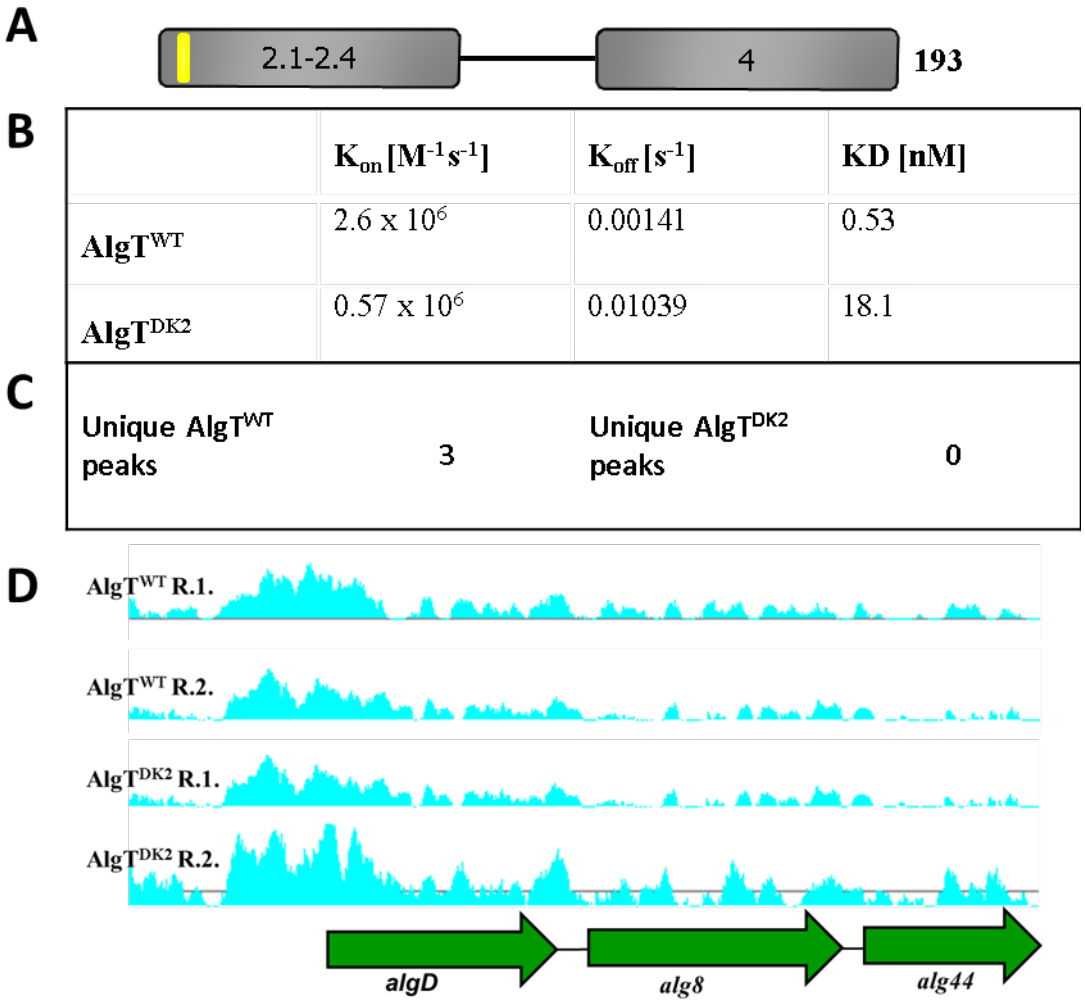
524 Zhang, Y. et al., 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9).

525 Zhou, Y.N., Walter, W. a & Gross, C. a, 1992. A mutant sigma 32 with a small deletion in
526 conserved region 3 of sigma has reduced affinity for core RNA polymerase. *Journal of*
527 *bacteriology*, 174(15), pp.5005–12.

528

529

Figures



	K_{on} [$M^{-1}s^{-1}$]	K_{off} [s^{-1}]	KD [nM]
$AlgT^{WT}$	2.6×10^6	0.00141	0.53
$AlgT^{DK2}$	0.57×10^6	0.01039	18.1

Unique $AlgT^{WT}$ peaks	3	Unique $AlgT^{DK2}$ peaks	0

Figure 1. Molecular effects of the $AlgT^{DK2}$ mutation. (A) Schematic illustration of the *P. aeruginosa* AlgT protein annotated with regions 2.1 - 2.4 and region 4. The $AlgT^{DK2}$ mutation is marked yellow in region 2.1. (B) SPR determined binding affinities between *P. aeruginosa* $AlgT^{WT}$ or $AlgT^{DK2}$ to *E. coli* core RNAP. (C) Number of unique genome wide binding events resulting from $AlgT^{WT}$ and $AlgT^{DK2}$ interactions at promoter sites. (D) Visual inspection of ChIP sequence data (in replicates) showing $AlgT^{WT}$ and $AlgT^{DK2}$ binding events at the at the *algD* promoter region.

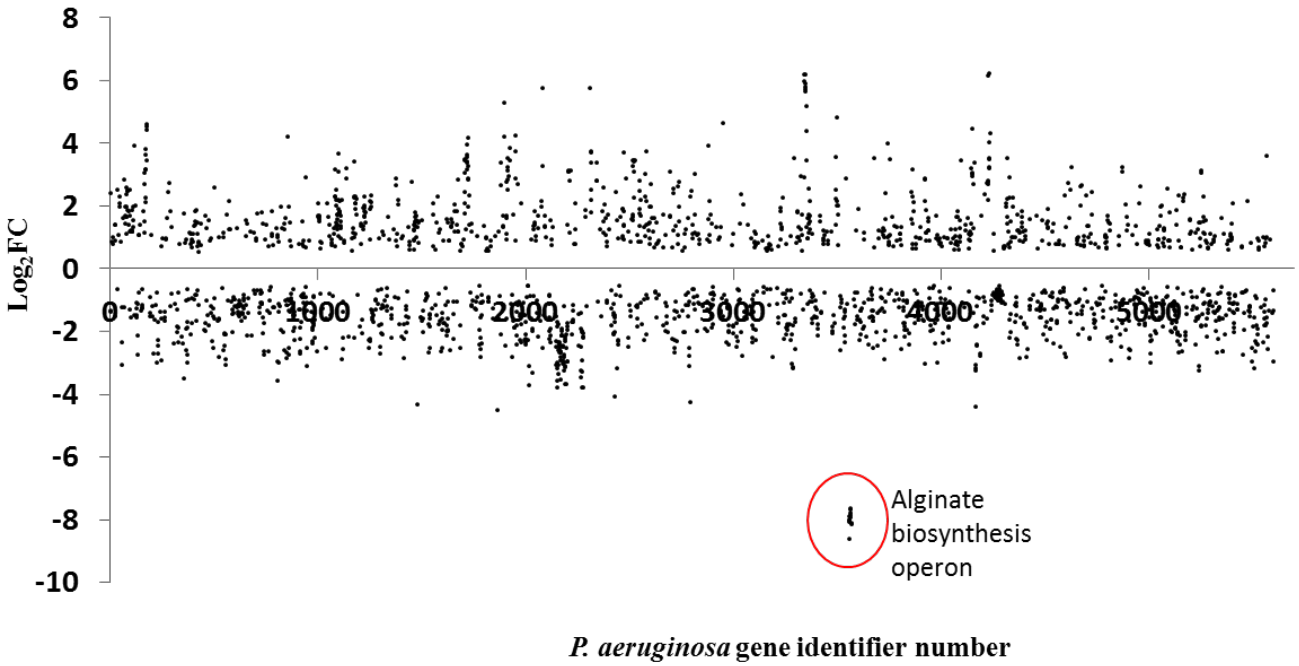
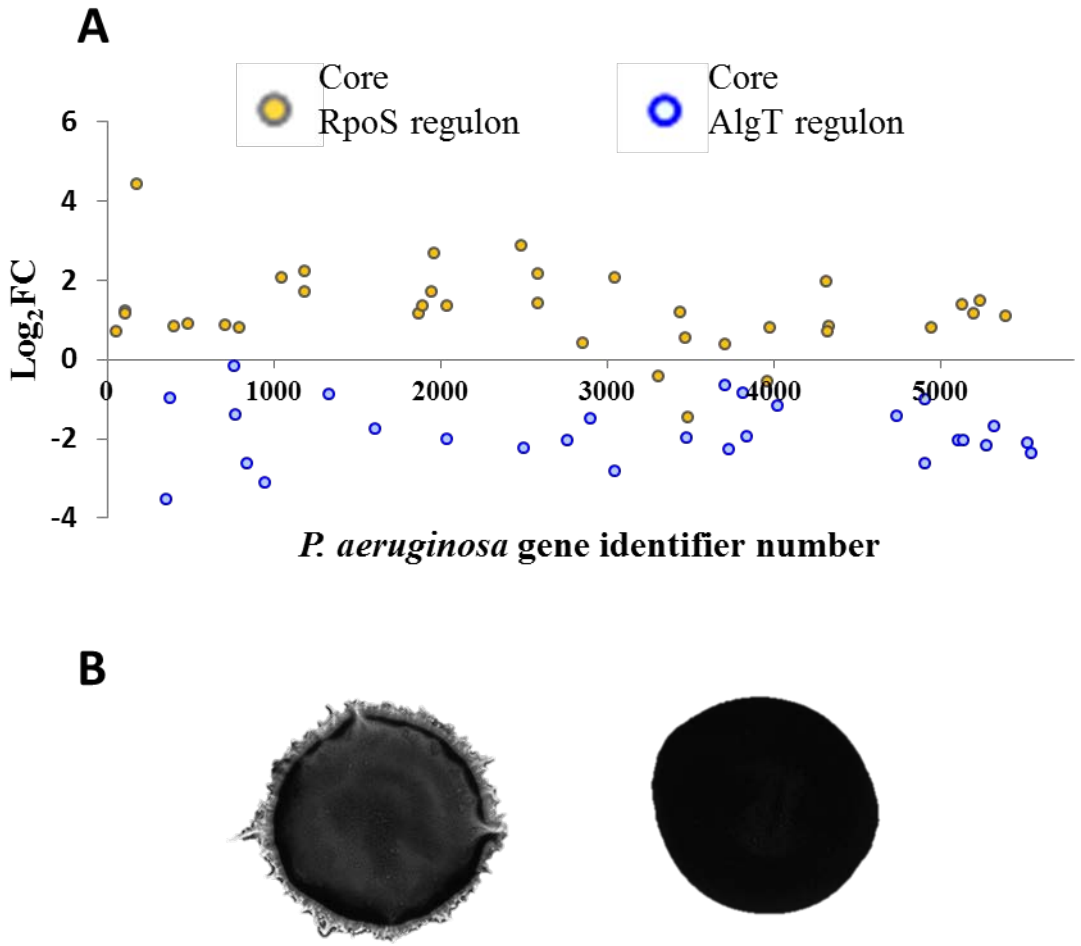


Figure 2. AlgT^{DK2} induced alterations on the gene expression profile. Gene expression profile of PAO1$^{\Delta mucA,\ algT(DK2)}$ expressing AlgT$^{(DK2)}$ compared to PAO1$^{\Delta mucA,algT(WT)}$ expressing AlgT$^{(WT)}$ after filtering for a false discovery rate of padj<0.05. PAO1 identifier numbers are mapped on the X axis, and the log_2FC change is mapped to the Y axis. The 12 gene operon involved in alginate biosynthesis is specifically marked on the graph.



548

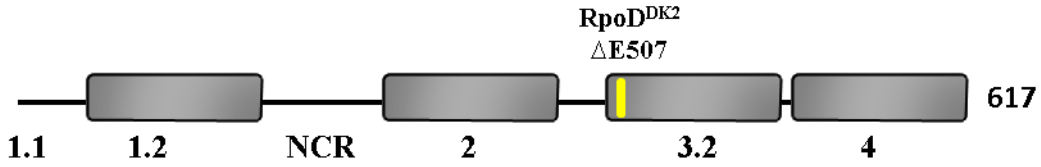
549 **Figure 3. Regulatory response of the AlgT and RpoS regulons from the AlgT^{DK2} mutation** (A)
550 Scatterplot showing the AlgT regulon (blue series) and RpoS regulon (yellow series) specific
551 regulatory response from the AlgT^{DK2} mutation. (B) Phenotype switching in response to AlgT^{DK2}
552 overexpression in PAO1$^{\Delta mucA,algT(DK2)}$. Shown to the left is non-induced PAO1$^{\Delta mucA,algT(DK2)}$
553 displaying a nonmucoid phenotype. Shown to the right is PAO1$^{\Delta mucA,algT(DK2)}$ overexpressing
554 AlgT^{DK2} which results in a phenotypic shift to a mucoid, alginate over-producing phenotype due to
555 restoration of AlgT^{DK2}-RNAP equilibrium.

556

557

558

A



B

	K_{on} [$M^{-1}s^{-1}$]	K_{off} [s^{-1}]	KD [nM]
RpoD^{WT}	1.7×10^6	0.0029	1.65
RpoD^{DK2}	3.1×10^6	0.0087	2.79

C

Unique peak signature for PAO1$^{rpoD(WT)}$ expressing RpoD^{WT}	3	Unique peak signature for PAO1$^{rpoD(DK2)}$ expressing RpoD^{DK2}	9

559

560 **Figure 4**. **Molecular effects of the RpoD^{DK2} mutation.** (A) Schematic representation of *P.*
561 *aeruginosa* RpoD with regions 1.1, 1.2, non-coding region (NCR), 2, 3.2, and region 4 annotated.
562 The RpoD^{DK2} ΔE507 deletion is marked yellow. (B) SPR determined affinities between *P.*
563 *aeruginosa* RpoD^{WT} and RpoD^{DK2} to *E. coli* core RNAP. (C) Number of unique genome wide
564 binding events resulting from RpoD^{WT} and RpoD^{DK2} interactions at promoter sites. Only 3 and 9
565 promoters were called as unique binding events for the RpoD^{WT} and RpoD^{DK2}, respectively,
566 indicating that the RpoD^{DK2} mutation does not affect the ability of the RpoD^{DK2}-RNAP to recognise
567 its promoter sequences.

568

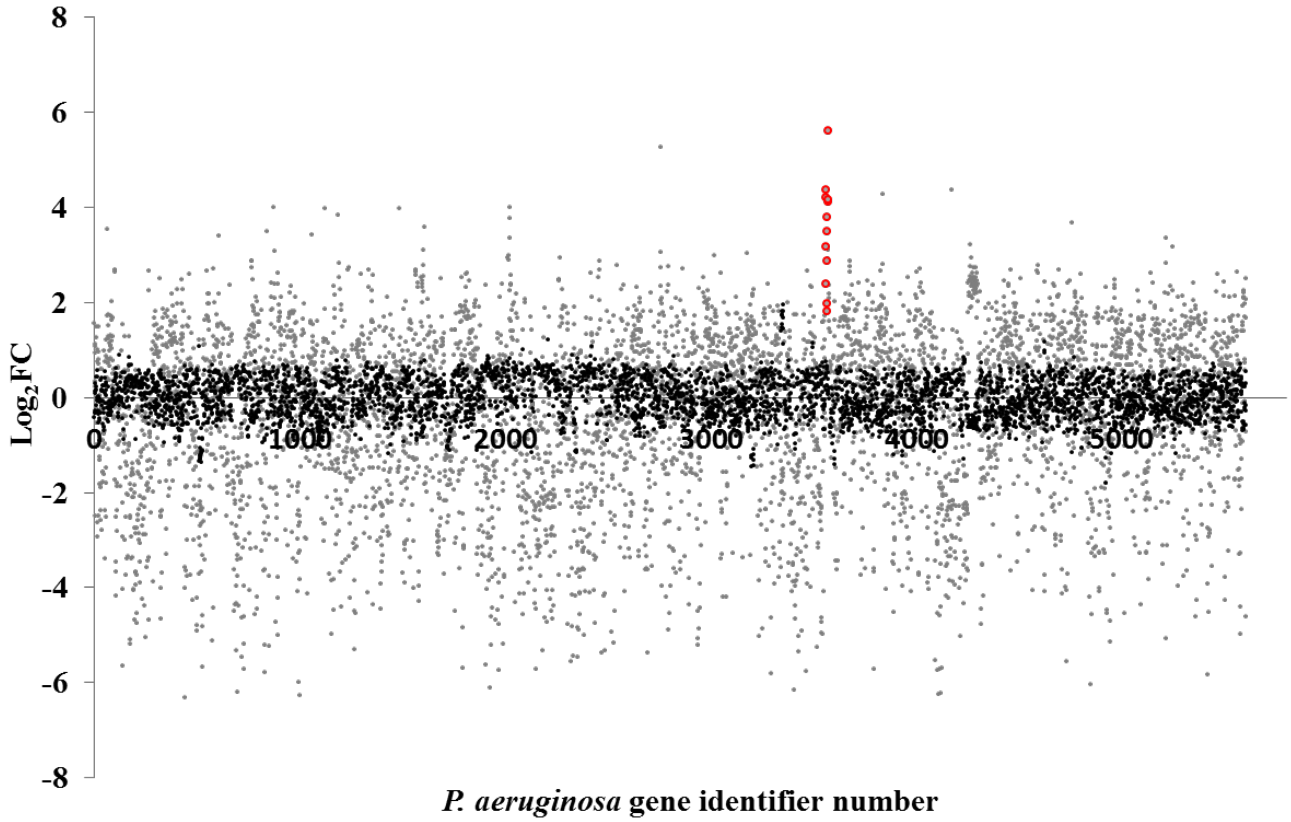


Figure 5. Gene expression profiles illustrating the epistatic effects generated by the PAO1$^{\Delta mucA,algT(DK2),rpoD(DK2)}$ mutations. The black series denotes the expression profile of PAO1$^{rpoD(DK2)}$ expressing RpoD^{DK2} compared to PAO1$^{rpoD(WT)}$ expressing RpoD^{WT}. The grey graph denotes the expression profile of PAO1$^{\Delta mucA,algT(DK2),rpoD(DK2)}$ expressing RpoD^{DK2} against PAO1$^{rpoD(DK2)}$ expressing RpoD^{DK2}. The red series denotes the specific gene expression values of the 12 gene operon responsible for alginate synthesis.

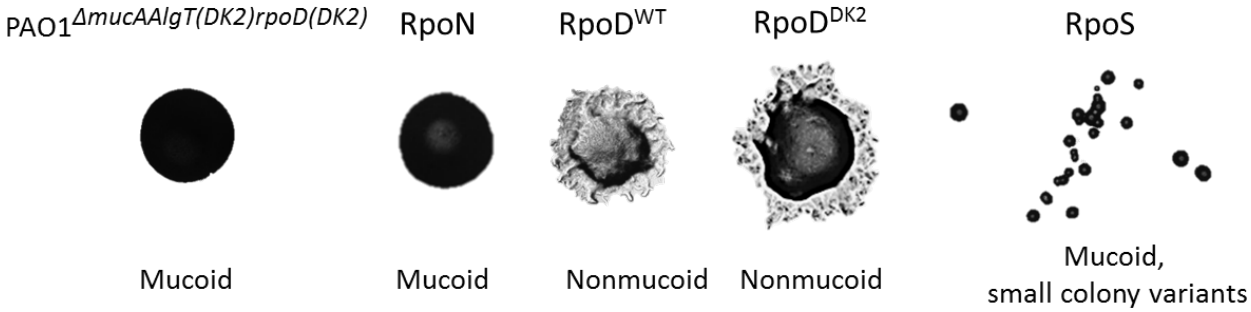


Figure 6. Robustness of the mucoid phenotype. Phenotype variations of PAO1$^{\Delta mucA,algT(DK2)rpoD(DK2)}$ due to *in vivo* overexpression of the sigma factors RpoN, RpoD^{WT}, RpoD^{DK2}, and RpoS. While overexpression of RpoN did not produce any phenotypic changes compared to PAO1$^{\Delta mucA,algT(DK2),rpoD(DK2)}$, overexpression of RpoS resulted in small colony variant which remained mucoid. Only by overexpression of either RpoD^{WT} or RpoD^{DK2} were we able to produce a phenotypic shift back to the nonmucoid phenotype.

589 **Table 1**

Strains used in this study	Features	Reference
PAO1		
PAO1$^{\Delta mucA}$		(Damkiaer et al. 2013)
PAO1$^{\Delta mucA,\ algT(DK2)}$		(Damkiaer et al. 2013)
PAO1$^{\Delta mucA,\ algT(DK2),\ rpoD(DK2)}$		(Damkiaer et al. 2013)
PAO1$^{rpoD(DK2)}$		(Damkiaer et al. 2013)
E. coli DH5-α		
E. coli Rosetta(DE3)		Novagen, Merck Millipore
Vectors used in this study		
pJN105		(Schulz et al. 2015)
pJN105-$algT^{(WT)}$		(Schulz et al. 2015)
pJN105-$algT^{(DK2)}$		This study
pJN105-$rpoD^{(WT)}$		(Schulz et al. 2015)
pJN105-$rpoD^{(DK2)}$		This study
pJN105-$rpoN^{(WT)}$		(Schulz et al. 2015)
pJN105-$rpoS^{(WT)}$		This study
pET28		Novagen, Merck Millipore
pET28-$algT^{(WT)}$		This study
pET28-$algT^{(DK2)}$		This study
pET28-$rpoD^{(WT)}$		This study
pET28-$rpoD^{DK2)}$		This study

590

591 **Table 2**

Primers used in this study	Sequence 5´-3´
ChIP-AlgT-F	CAGCGCGGAGACGAGCGGGCTTTCG
ChIP-AlgT-R	CGAAAGCCCGCTCGTCTCCGCGCTG
ChIP-RpoD-F	GCAAGGTACTGAAGATCGCCAAACCGATCTCCATG
ChIP-RpoD-R	CATGGAGATCGGTTTGGCGATCTTCAGTACCTTGC
rpoSF	ACTGGAATTCTTAACTTTAAGGAGGAGATATAATGGCACTCAAAAAAGAAGGGC
rpoSR	ACTGTCTAGATCACTGGAACAGCGCGTCA
AlgT_F	AGATACATATGCTAACCCAGGAACAGGAT
AlgT_R	AGGTAAAGCTTCAGGCTTCTCGCAACAAAG
RpoD_F	AAAGCCATATGTCCGGAAAAGCGCAACA
RpoD_R	CGCGAAGCTTCACTCGTCGAGGAAGGAGC

592

Supplementary material

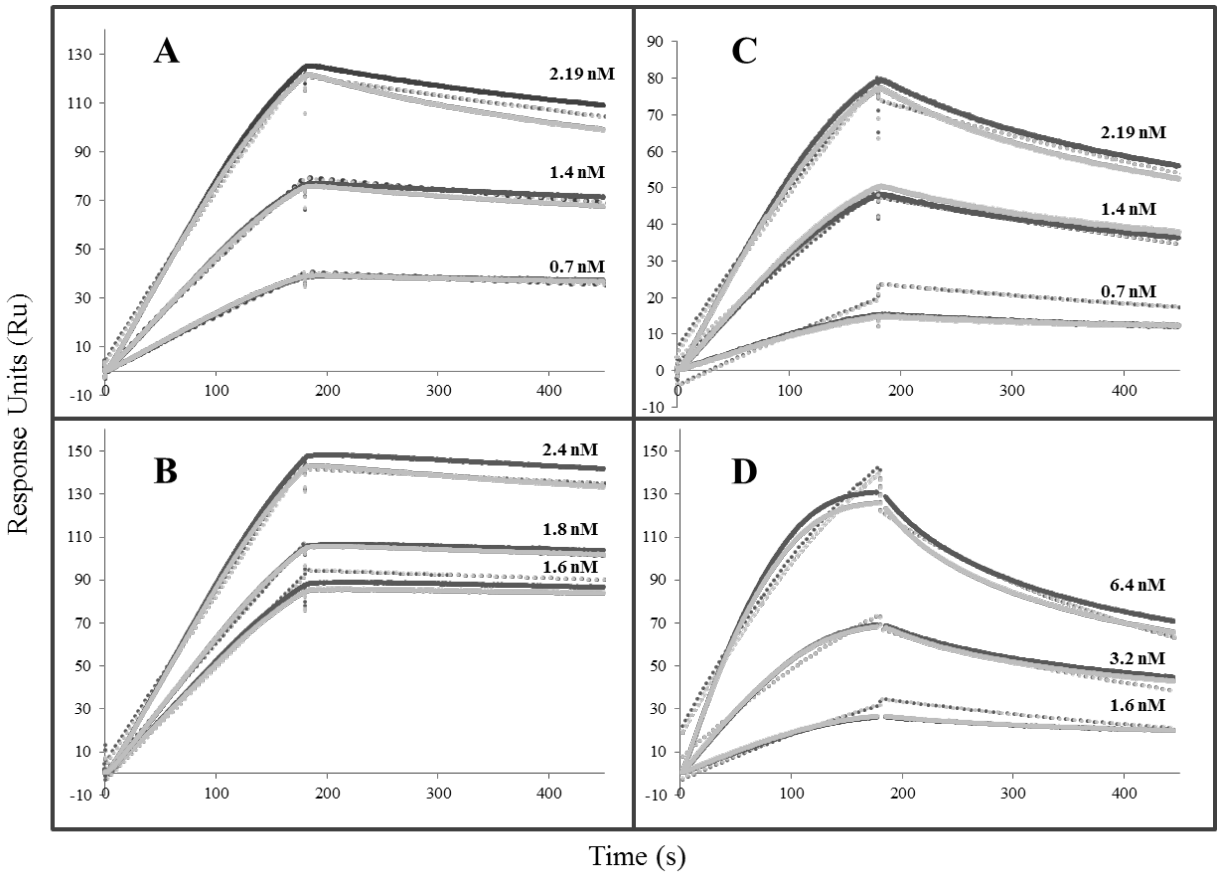


Figure S1. Sensorgrams of sigma factor interactions to *E. coli* core RNAP, measured by SPR. S1.A: Immobilised $AlgT^{WT}$ binding to *E. coli* core RNAP. S1.B: Immobilised $AlgT^{DK2}$ binding to *E. coli* core RNAP. S1.C. Immobilised $RpoD^{WT}$ binding to *E. coli* core RNAP. S1.D. Immobilised $RpoD^{DK2}$ binding to *E. coli* core RNAP.